V.I. SMIRNOV

A COURSE OF HIGHER MATHEMATICS

2

ADVANCED CALCULUS

A COURSE OF
Higher Mathematics

VOLUME II

V. I. SMIRNOV

Translated by
D. E. BROWN

Translation edited by
I. N. SNEDDON
Simson Professor in Mathematics
University of Glasgow

PERGAMON PRESS
OXFORD · LONDON · EDINBURGH · NEW YORK
PARIS · FRANKFURT

ADDISON-WESLEY PUBLISHING COMPANY, INC.
READING, MASSACHUSETTS · PALO ALTO · LONDON
1964

INTRODUCTION

SOME account of the history of this five-volume course of higher mathematics has been given in the Introduction to Vol. I of the present English edition.

The first Russian edition of the present volume appeared (in 1926) under the joint authorship of Professor Smirnov and the late Professor J. D. Tamarkin but later editions, prepared after Professor Tamarkin had settled in the U.S.A. and consisting of a drastic revision of (and many additions to) the original material, contained only Professor Smirnov's name. This volume is made up of a course of advanced calculus which is of great use to students of mathematics and which provides the physicists and engineers with a complete set of those tools, based upon the theory of functions of real variables, which are indispensable for the study of the classical branches of mathematical physics.

It consists essentially of five distinct parts, although there are strong links connecting all of them. There is a full discussion of the solution of ordinary differential equations with many applications to the treatment of physical problems. This is followed by an account of the properties of multiple integrals and of line integrals, with a valuable section on the theory of measurable sets and of multiple integrals.

The mathematics necessary to the discussion of problems in classical field theories is discussed in a section on vector algebra and vector analysis; the methods developed are illustrated not only by applications to physics but also by an account of the elements of differential geometry in three-dimensional space.

After this there comes an elementary but full account of Fourier series.

The principles and techniques developed in these sections are then applied to the discussion of the solution of the partial differential equations of classical mathematical physics.

The clarity of Prof. Smirnov's exposition and the width of his knowledge of the mathematical techniques effective in the study of the physical sciences makes the whole course a most valuable one for

the student anxious not only to learn the methods of advanced calculus but also to understand the influences which have motivated their development.

I. N. SNEDDON

PREFACE TO THE SIXTH EDITION

THIS edition of the second volume differs considerably from the previous one. The first chapter of the previous edition, containing the theory of complex numbers, the principles of higher algebra, and integration of functions, was transferred to the first volume. At the same time, all material referring to the principles of vector algebra was taken from Volume I to Volume II. This material was incorporated in Chapter IV, together with vector analysis

The presentation of the remaining chapters underwent substantial changes. This refers particularly to Chapters III, VI and VII. A special paragraph containing the theory of dimensions and the rigorous theory of multiple integrals was added to Chapter III. A certain re-distribution of material was carried out in Chapter VI, and a proof was added of the closure equation on the basis of Weierstrass' theorem on polynomial approximation to continuous functions. Chapter VII now contains additional material on the propagation of spherical and cylindrical waves and Kirchhoff's formula for the solution of the wave equation. The explanation of linear differential equations with constant coefficients is introduced at first without using the symbolic method.

First paragraphs of each chapter have retained their explanatory character. The book is arranged in such a way that the basic material in larger type can be studied without reference to the examples or complementary theoretical material printed in small type.

I should like to express my deep gratitude to Prof. Fikhtengol'ts, who has read the manuscript of this edition, for his valuable suggestions concerning the style and arrangement of the book.

V. SMIRNOV

PREFACE TO THE FOURTEENTH EDITION

THE GENERAL arrangement of the present edition is the same as that of the previous edition. However, small alterations were introduced in many places with the aim of clarifying the style and achieving greater readability.

Most substantial changes were carried out in Paragraph 9 (Chapter III), "Supplementary remarks on the theory of multiple integrals".

In Chapter VII, devoted to simple problems of mathematical physics, the formulation of conditions for the solution of a series of basic problems was clarified. References to matters explained in detail in Volume IV have been added in several places in Chapter VII.

V. SMIRNOV

CHAPTER I

ORDINARY DIFFERENTIAL EQUATIONS

§ 1. Equations of the first order

1. General principles. A differential equation is defined as an equation which contains, in addition to independent variables and unknown functions, derivatives or differentials of the unknown functions [1, 51]. If the functions appearing in a differential equation depend on a *single independent variable*, the equation is called an *ordinary differential* equation. On the other hand, if partial derivatives of the functions with respect to certain of the independent variables appear in the equation, it is called a *partial differential equation.* We confine ourselves to ordinary differential equations in the present chapter, the greater part of which is devoted to the case of a single equation containing one unknown function.

Let x be the independent variable, and y the required function of this variable. The general form of the differential equation becomes:

$$\Phi(x, y, y' y'', \ldots, y^{(n)}) = 0.$$

The *order of the differential equation* is defined as the order n of the highest order derivative of the function that appears in the equation. We shall consider ordinary differential equations of the *first order* in the present article. The general form of this equation is:

$$\Phi(x, y, y') = 0 \tag{1}$$

or, on solving with respect to y':

$$y' = f(x, y). \tag{2}$$

If *a function*

$$y = \varphi(x) \tag{3}$$

satisfies the differential equation, i.e. if the equation reduces to an identity on replacing y and y' by $\varphi(x)$ and $\varphi'(x)$, *the function* $\varphi(x)$ *is said to be a solution of the differential equation.*

The problem of finding a solution of a differential equation is alternatively referred to as the task of integrating the equation.

If x and y are considered as the coordinates of points on a plane, differential equation (1) [or (2)] expresses a relationship between coordinates of points on a certain curve and the slopes of the tangents to the curve at these points. A curve corresponds to the solution (3) of the equation, the points and tangential slopes of which satisfy the equation. This curve is referred to as *an integral curve of the given differential equation.*

In the simplest case, when the right-hand side of equation (2) does not contain y, a differential equation is obtained of the form:

$$y' = f(x).$$

Finding the solutions of this equation is the primary task of the integral calculus [I, 86], and the total set of solutions is given by the formula:

$$y = \int f(x)\,\mathrm{d}x + C,$$

where C is an arbitrary constant. We thus obtain in this elementary case a solution of the differential equation containing an arbitrary constant. We shall see that a solution containing an arbitrary constant is also obtained in the general case of a first order differential equation; such a solution is referred to as the *general solution of the equation.* On assigning different numerical values to the arbitrary constant, we obtain the various so-called *particular solutions of the equation.*

We give in the following sections some particular types of first order equation, integration of which leads to evaluation of indefinite integrals — or, as it may sometimes be expressed, *their integration reduces to quadrature.*†

2. Equations with separable variables. On replacing y' in equation (2) by the quotient $\mathrm{d}y/\mathrm{d}x$, multiplying both sides by $\mathrm{d}x$, and carrying all terms to the left-hand side, we can write (2) in the form:

$$M(x, y)\,\mathrm{d}x + N(x, y)\,\mathrm{d}y = 0, \tag{4}$$

which will be more convenient in some cases. Both variables x and y play an identical role here in the equation, so that (4) does not bind us to the choice of unknown function: we can take either x or y for this, as we wish.

† Evaluation of an integral has a direct connection with evaluation of an area, whence the term "quadrature".

We assume that each of the functions $M(x, y)$ and $N(x, y)$ consists of the product of two factors, one of which depends only on x, and the other only on y:

$$M_1(x)\,M_2(y)\,\mathrm{d}x + N_1(x)\,N_2(y)\,\mathrm{d}y = 0. \tag{5}$$

On dividing both sides of the equation by $M_2(y)\,N_1(x)$, we reduce it to the form:

$$\frac{M_1(x)}{N_1(x)}\,\mathrm{d}x + \frac{N_2(y)}{M_2(y)}\,\mathrm{d}y = 0, \tag{6}$$

so that the coefficient of $\mathrm{d}x$ now depends only on x, and the coefficient of $\mathrm{d}y$ only on y. Equation (5) is called an equation with separable variables [1, 93], whilst the method itself of reduction to the form (6) is called separation of the variables.

The left-hand side of equation (6) is the differential of the following expression:

$$\int \frac{M_1(x)}{N_1(x)}\,\mathrm{d}x + \int \frac{N_2(y)}{M_2(y)}\,\mathrm{d}y,$$

and the equating to zero of the differential of this expression means that the expression itself is equal to an arbitrary constant:

$$\int \frac{M_1(x)}{N_1(x)}\,\mathrm{d}x + \int \frac{N_2(y)}{M_2(y)}\,\mathrm{d}y = C, \tag{7}$$

where C is the arbitrary constant. This formula gives an infinite se of solutions and, from the geometrical point of view, is the implici equation of a family of integral curves. If the quadratures are carried out in (7) and we solve the equation with respect to y, we obtain the explicit equation of the family of integral curves (the solution of the differential equation):

$$y = \varphi(x, C).$$

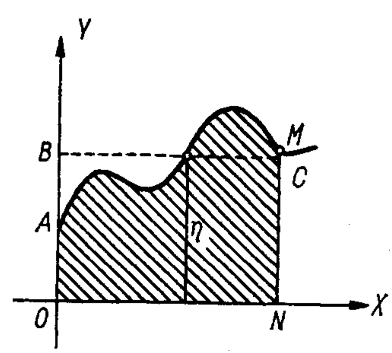

FIG. 1

Example. The area $OAMN$, bounded by the coordinate axes, the segment AM of a curve and its ordinate MN (Fig. 1), is equal in magnitude to a rectangular area $OBCN$ with the same base $ON = x$ and with height η:

$$\int_0^x y\,\mathrm{d}x = x\eta; \quad \eta = \frac{1}{x}\int_0^x y\,\mathrm{d}x. \tag{8}$$

The magnitude η is called the *average ordinate of the curve in the interval* $(0, x)$.

Let us find the curves whose average ordinates are proportional to the extreme ordinate NM. We have on the basis of (8):

$$\int_0^x y \, dx = kxy, \tag{9}$$

where k is the coefficient of proportionality. On differentiating both sides of equation (9), we get the differential equation:

$$y = ky + kxy', \text{ or } xy' = ay, \tag{10}$$

where

$$a = \frac{1-k}{k}. \tag{11}$$

Unwanted solutions may have been introduced on differentiation, since the equality of the derivatives implies only that the functions themselves differ

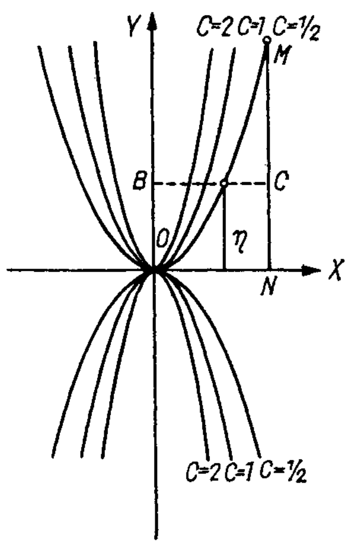

FIG. 2

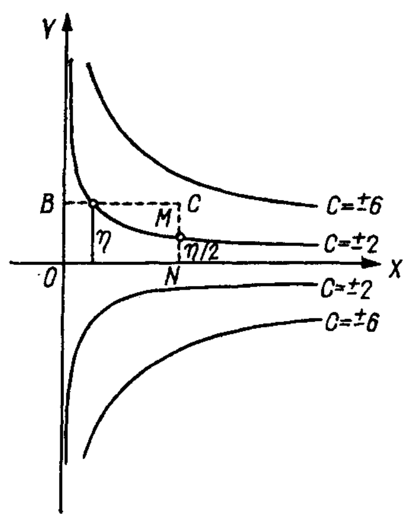

FIG. 3

by a constant. There are no unwanted solutions in the present case, however. It follows from equation (10), obtained by differentiation of (9), that both sides of (9) can only differ by a constant; but it is immediately evident that both these sides are zero for $x = 0$, so that the constant in question is zero, i.e. every solution of (10) is also a solution of (9). We pass to the integration of (10). It can be written as:

$$x \frac{dy}{dx} = ay,$$

and the variables can be separated:

$$\frac{dy}{y} = a \frac{dx}{x}.$$

We obtain on integration:

$$\log y = a \log x + C_1 \text{ or } y = Cx^a, \tag{12}$$

where $C = e^{C_1}$ is an arbitrary constant.

From (11), as k increases from 0 to $+\infty$, a decreases from $+\infty$ to -1, and we must therefore take $a > -1$, so that the integral on the left-hand side of (9) never becomes meaningless. We have $a = 0$ for $k = 1$, and (12) gives as the obvious solution a family of straight lines parallel to axis OX. We have $a = 2$ for $k = 1/3$, which gives the family of parabolas (Fig. 2)

$$y = Cx^2,$$

for which the average ordinate is equal to a third of the extreme ordinate. With $k = 2$, we get the family of curves:

$$y = \frac{C}{\sqrt{x}},$$

for which the average ordinate is twice the extreme ordinate (Fig. 3).

3. Homogeneous equations. *A homogeneous equation* is defined as an equation of the form:

$$y' = f\left(\frac{y}{x}\right)^{\dagger}. \tag{13}$$

We preserve the previous independent variable x but introduce a new function u instead of y:

$$y = xu, \text{ whence } y' = u + xu'. \tag{14}$$

We transform our equation:

$$u + xu' = f(u) \text{ or } x\frac{du}{dx} = f(u) - u.$$

Separation of the variables gives:

$$\frac{dx}{x} + \frac{du}{u - f(u)} = 0.$$

We obtain, on denoting the coefficient of du by $\psi_1(u)$:

$$\log x + \int \psi_1(u)\, du = C_1,$$

whence

$$x = Ce^{-\int \psi_1(u) du} \text{ or } x = C\psi(u),$$

where $C = e^{C_1}$ is an arbitrary constant.

† We remark that the function $\varphi(x, y)$ of two variables is a function simply of the ratio y/x when, and only when, the magnitude of $\varphi(x, y)$ is unchanged on multiplying x and y by an arbitrary factor t, i.e. $\varphi(tx, ty) = \varphi(x, y)$. This condition is equivalent to $\varphi(x, y)$ being a homogeneous function of x and y of zero degree [1, 151].

On returning to the previous variable, we can write the equation of the family of integral curves in the form:

$$x = C\psi\left(\frac{y}{x}\right). \tag{15}$$

We consider a transformation of similitude in the plane XOY with centre of similitude at the origin. The transformation amounts to the point (x, y) being transferred to the new position:

$$x_1 = kx; \quad y_1 = ky \quad (k > 0) \tag{16}$$

or, which comes to the same thing, it amounts to multiplying the length of the radius vector to every point of the plane by k whilst preserving the direction. If M is the original position of a point, and M_1 the position of the same point after transformation, we have (Fig. 4):

$$\overline{OM_1} : \overline{OM} = x_1 : x = y_1 : y = k.$$

On applying transformation (16) to equation (15), we get the equation:

$$x_1 = kC\psi\left(\frac{y_1}{x_1}\right), \tag{17}$$

which does not differ from equation (15), in view of the arbitrariness of the constant C, i.e. transformation (16) does not alter the total set of curves (15) but only moves one curve of family (15) to the position of another curve of the same family. Any curve of family (15) can evidently be obtained from one definite curve of the family by using transformation (16), with appropriate choice of the constant k. The result obtained can be expressed as follows: *all the integral curves of a homogeneous equation can be obtained from one integral curve by means of the transformation of similitude, with centre of similitude at the origin.*

Equation (13) can be written as:

$$\tan \alpha = f(\tan \theta),$$

where $\tan \alpha$ is the slope of the tangent, and θ is the angle formed by the radius vector from the origin with the positive direction of axis OX. Equation (13) thus establishes a connection between angles α and θ, such that *the tangents to the integral curves of a homogeneous equation, drawn at the points of intersection of the curves with a straight line through the origin, must be mutually parallel* (Fig. 4).

It follows obviously from this property of the tangents that the transformation of similitude with centre of similitude at the origin

transforms one integral curve to another integral curve, since, on increasing the lengths of radius vectors of points of the curve in the same ratio, the directions of the tangents at the ends of the radius vectors are unchanged (Fig. 5).

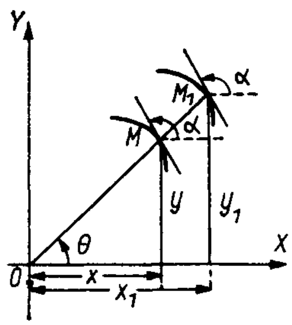

FIG. 4

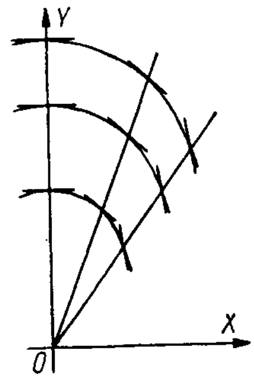

FIG. 5

If we apply the above transformation to the integral curve consisting of a straight line passing through the origin, we get the same line after transformation, so that in this case the above method of obtaining the set of integral curves from one of them fails.

Example. To find the curves such that the length MT of the tangent from its point of contact to its intersection with OX is equal to the length OT along OX (Fig. 6).

The equation of the tangent has the form:

$$Y - y = y' (X - x),$$

where (X, Y) are the current coordinates of the tangent. We find the intercept of the tangent on OX by putting $Y = 0$:

$$\overline{OT} = x - \frac{y}{y'},$$

and by hypothesis, $\overline{MT}^2 = \overline{OT}^2$, which gives us [I, 77]:

$$\frac{y^2}{y'^2} + y^2 = \left(x - \frac{y}{y'}\right)^2,$$

whence we obtain the differential equation:

$$y' = \frac{2xy}{x^2 - y^2}, \tag{18}$$

which evidently belongs to the homogeneous type.

We introduce a new function u instead of y, in accordance with the formula:

$$y = xu; \quad y' = xu' + u.$$

We have, on substituting in equation (18):

$$xu' + u = \frac{2u}{1-u^2} \qquad \text{or} \qquad x\frac{\mathrm{d}u}{\mathrm{d}x} - \frac{u+u^3}{1-u^2} = 0. \tag{19}$$

which gives, on separating the variables:

$$\frac{\mathrm{d}x}{x} - \frac{1-u^2}{u+u^3}\,\mathrm{d}u = 0. \tag{20}$$

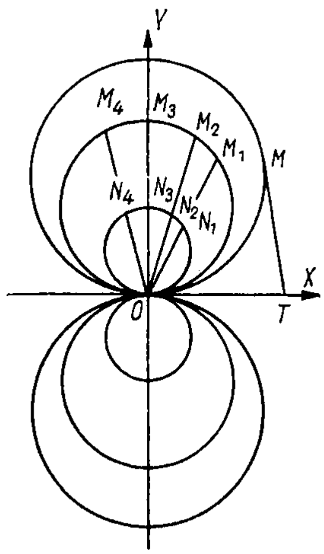

FIG. 6

Integration gives us:

$$\frac{x(u^2+1)}{u} = C,$$

or, on returning to the previous variable y:

$$x^2 + y^2 - Cy = 0, \tag{21}$$

i.e. the required curves are circles passing through the origin and touching the axis OX at this point (Fig. 6).

We divided both sides of equation (19) by $(u + u^3)$ in order to pass from (19) to (20), and we might have lost the solution $u = 0$, or, what comes to the same thing, $y = 0$. We see on substituting in equation (18) that this is in fact a solution. The solution is contained in equation (21), however; we can obtain it by dividing both sides of (21) by C, then setting $C = \infty$.

Each circle of family (21) can be got from any one of them by the transformation of similitude with centre of similitude at the origin, so that (Fig. 6):

$$\frac{\overline{OM_1}}{\overline{ON_1}} = \frac{\overline{OM_2}}{\overline{ON_2}} = \frac{\overline{OM_3}}{\overline{ON_3}} = \cdots$$

The differential equation:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f\left(\frac{ax+by+c}{a_1x+b_1y+c_1}\right), \tag{22}$$

reduces to a homogeneous form, as we shall now show. We introduce new variables ξ and η in place of x and y:

$$x = \xi + \alpha; \qquad y = \eta + \beta, \tag{23}$$

where α and β are constants which we proceed to define.

Equation (22) becomes in the new variables:

$$\frac{\mathrm{d}\eta}{\mathrm{d}\xi} = f\left(\frac{a\xi + b\eta + a\alpha + b\beta + c}{a_1\xi + b_1\eta + a_1\alpha + b_1\beta + c_1}\right).$$

We define α and β by the conditions:

$$a\alpha + b\beta + c = 0\,, \quad a_1\alpha + b_1\beta + c_1 = 0\,.$$

The equation now reduces to the homogeneous form:

$$\frac{d\eta}{d\xi} = f\left[\frac{a + b\,\dfrac{\eta}{\xi}}{a_1 + b_1\,\dfrac{\eta}{\xi}}\right].$$

Transformation (23) corresponds to parallel displacements of the axes, with the origin becoming the point of intersection (α, β) of the straight lines

$$ax + by + c = 0 \quad \text{and} \quad a_1x + b_1y + c_1 = 0. \tag{24}$$

The results previously obtained will thus be applicable to equation (22), with the only difference that the role of origin will be played by the **point (α, β).**

If the straight lines (24) are parallel, the transformation mentioned cannot be carried out. But in this case, as we know from analytic geometry, the coefficients in equations (24) must be proportional:

$$\frac{a_1}{a} = \frac{b_1}{b}\,\lambda \text{ and } a_1x + b_1y = \lambda\,(ax + by);$$

on introducing a new variable u instead of y:

$$u = ax + by,$$

it can easily be seen that we obtain an equation with separable variables.

We shall encounter below an extremely important application of homogeneous equations, to the investigation of fluid flow.

4. Linear equations; Bernoulli's equation. An equation of the form:

$$y' + P(x)\,y + Q(x) = 0\,. \tag{25}$$

is called a linear equation of the first order.

We start by considering the *equation with no term $Q(x)$*:

$$z' + P(x)\,z = 0\,.$$

The variables are separable here:

$$\frac{dz}{z} + P(x)\,dx = 0\,,$$

and we get:

$$z = Ce^{-\int P(x)\,dx}\,. \tag{26}$$

We integrate the given linear equation (25) by using the method of varying the arbitrary constant, i.e. we seek a solution of the equation in a form analogous to the form (26) for z:

$$y = u e^{-\int P(x)\,dx}, \tag{27}$$

where u is no longer a constant, but the required function of x. We get by differentiation:

$$y' = u' e^{-\int P(x)\,dx} - P(x)\, u e^{-\int P(x)\,dx}.$$

Substitution in equation (25) gives:

$$u' e^{-\int P(x)\,dx} + Q(x) = 0$$

$$u' = -Q(x)\, e^{\int P(x)\,dx}, \text{ whence } u = C - \int Q(x)\, e^{\int P(x)\,dx}\, dx.$$

We finally get, by equation (27) for y:

$$y = e^{-\int P(x)\,dx} \left[C - \int Q(x)\, e^{\int P(x)\,dx}\, dx\right]. \tag{28}$$

When determining y by this formula, we only need to take one each of the values of the indefinite integrals

$$\int P(x)\,dx \text{ and } \int Q(x)\, e^{\int P(x)\,dx}\, dx,$$

since the addition of arbitrary constants to these only changes the value of C.

If we replace them by definite integrals with variable upper limits [I, 96], we can re-write (28) as:

$$y = e^{-\int_{x_0}^{x} P(x)\,dx} \left[C - \int_{x_0}^{x} Q(x)\, e^{\int_{x_0}^{x} P(x)\,dx}\, dx\right], \tag{29}$$

where x_0 is a definite number, though chosen arbitrarily. On substituting the value $x = x_0$ for the variable upper limit, the right-hand side of the formula written is equal to C, since integrals with identical upper and lower limits are equal to zero; in other words, the constant C in formula (29) is the value of the function y at $x = x_0$. This value, which we denote by y_0, is called the *initial value of the solution*.

We denote this fact by writing:

$$y\,|_{x=x_0} = y_0. \tag{30}$$

If the initial value of the required solution is given for $x = x_0$, (29) yields a completely defined solution of the equation:

$$y = \mathrm{e}^{-\int\limits_{x_0}^{x} P(x)\,\mathrm{d}x} \left[y_0 - \int\limits_{x_0}^{x} Q(x)\, \mathrm{e}^{\int\limits_{x_0}^{x} P(x)\,\mathrm{d}x}\, \mathrm{d}x \right]. \tag{31}$$

Condition (30) is called the initial condition and is equivalent geometrically to the integral curve being sought which passes through the given point (x_0, y_0).

If we take $Q(x) \equiv 0$, we obtain the solution of the homogeneous equation

$$y' + P(x)\, y = 0.$$

satisfying condition (30):

$$y = y_0\, \mathrm{e}^{-\int\limits_{x_0}^{x} P(x)\,\mathrm{d}x}. \tag{31_1}$$

It follows from (29) that solutions of a linear differential equation have the form:

$$y = \varphi_1(x)\, C + \varphi_2(x), \tag{32}$$

i.e. *y is a linear function of the arbitrary constant.*

Let y_1 be a solution of equation (25). On setting

$$y = y_1 + z,$$

we get the equation for z:

$$z' + P(x)\, z + [y_1' + P(x)\, y_1 + Q(x)] = 0.$$

The sum appearing in square brackets is equal to zero, since y_1 is a solution of equation (25) by hypothesis. It follows that z is a solution of the equation when the term $Q(x)$ is absent and is defined by (26), whence:

$$y = y_1 + C\mathrm{e}^{-\int P(x)\,\mathrm{d}x}. \tag{33}$$

We now assume that a further solution y_2 is known of equation (25), and we let this solution be obtained from (33) with $C = a$:

$$y_2 = y_1 + a\mathrm{e}^{-\int P(x)\,\mathrm{d}x}. \tag{34}$$

If we eliminate $\mathrm{e}^{-\int P(x)\,\mathrm{d}x}$ from (33) and (34), we obtain an expression for the solution of a linear equation in terms of two of its solutions y_1 and y_2:

$$y = y_1 + C_1(y_2 - y_1), \tag{35}$$

where C_1 is an arbitrary constant replacing C/a in the previous notation. The following relationship follows from (35):

$$\frac{y_2 - y}{y - y_1} = \frac{1 - C_1}{C_1} = C_2, \tag{36}$$

which shows that the ratio $(y_2-y)/(y-y_1)$ is a constant, i.e. *the family of integral curves of a linear equation is a family of curves that divide the segment of ordinate between any two curves of the family in a constant ratio.*

If two integral curves L_1 and L_2 of a linear equation are known, any other integral curve L is defined by the constant value of the ratio (Fig. 7)

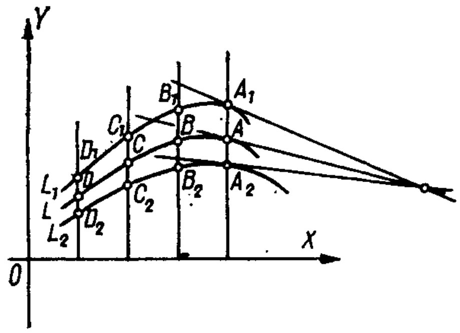

Fig. 7

$$\frac{\overline{AA_2}}{\overline{A_1A}} = \frac{\overline{BB_2}}{\overline{B_1B}} = \frac{\overline{CC_2}}{\overline{C_1C}} = \frac{\overline{DD_2}}{\overline{D_1D}} = \dots .$$

It follows from this equation that chords A_1B_1, AB and A_2B_2 must either meet in a single point or be parallel. On letting $\overline{B_1B_2}$ approach indefinitely near to $\overline{A_1A_2}$, the directions of these chords become the directions of the tangents to the curves at A_1, A and A_2, and we obtain the following property of the tangents to the integral curves of a linear equation: *the tangents to the integral curves of a linear equation at the points of intersection of the curves with a line drawn parallel to OY either intersect in a single point or are parallel.*

Examples. **1.** We consider the transient current in a circuit with self-inductance. Let i be the current, v the voltage, R the resistance in circuit, and L the self-inductance.

The following relationship is valid:

$$v = Ri + L\frac{di}{dt},$$

whence we obtain the linear equation for i:

$$\frac{di}{dt} + \frac{R}{L}i - \frac{v}{L} = 0.$$

We take R and L as constants and v as a given function of time t, and evaluate the integrals appearing in formula (31):

$$\int_0^t P dt = \int_0^t \frac{R}{L} dt = \frac{R}{L} t; \qquad \int_0^t Q e^{\int_0^t P dt} dt = -\frac{1}{L}\int_0^t v e^{\frac{R}{L}t} dt.$$

If we let i_0 denote the initial value of i, i.e. the value of the current at $t = 0$, (31) gives us the following formula for determining i at any required instant:

$$i = e^{-\frac{R}{L}t}\left(i_0 + \frac{1}{L}\int_0^t ve^{\frac{R}{L}t}\,dt\right).$$

We have in the case of constant voltage v:

$$i = \left(i_0 - \frac{v}{R}\right)e^{-\frac{R}{L}t} + \frac{v}{R}.$$

The factor $e^{-Rt/L}$ rapidly decreases as t increases, and in practice the process an be assumed to have reached the steady state after a short space of time, the urrent being then given by Ohm's law: $i = v/R$.

In the particular case of $i_0 = 0$ we get the formula:

$$i = \frac{v}{R}\left(1 - e^{-\frac{R}{L}t}\right) \tag{37}$$

for the current in a *closed* circuit.

The constant L/R is called the *time constant* of the circuit.

We consider a voltage v of sinusoidal form, $v = A \sin \omega t$. We obtain by using (31):

$$i = e^{-\frac{R}{L}t}\left[i_0 + \frac{A}{L}\int_0^t e^{\frac{R}{L}t} \sin \omega t\,dt\right].$$

It is easily seen [I, 201] that:

$$\int e^{\frac{R}{L}t} \sin \omega t\,dt = e^{\frac{R}{L}t}\left[\frac{RL}{\omega^2 L^2 + R^2} \sin \omega t - \frac{\omega L^2}{\omega^2 L^2 + R^2} \cos \omega t\right]$$

and therefore:

$$\int_0^t e^{\frac{R}{L}t} \sin \omega t\,dt = e^{\frac{R}{L}t}\left[\frac{RL}{\omega^2 L^2 + R^2} \sin \omega t - \frac{\omega L^2}{\omega^2 L^2 + R^2} \cos \omega t\right] + \frac{\omega L^2}{\omega^2 L^2 + R^2}.$$

We obtain on substituting in the expression for i:

$$i = \left(i_0 + \frac{\omega LA}{\omega^2 L^2 + R^2}\right) e^{-\frac{R}{L}t} + \frac{RA}{\omega^2 L^2 + R^2} \sin \omega t - \frac{\omega LA}{\omega^2 L^2 + R^2} \cos \omega t. \tag{38}$$

The first term, containing the factor $e^{-Rt/L}$, is rapidly damped, and in practice the current will be given in a short space of time after $t = 0$ by the sum of the two remaining terms of (38). This sum consists of a sinusoidal quantity of the same frequency ω as the voltage v, but with different amplitude and phase. We also notice that the sum giving the steady-state current does not depend on the initial value of the current i_0.

2. The resistance R cannot be reckoned as a constant in switching processes, when a spark appears. It increases from an initial value R_0 to infinity (at the instant τ of breaking contact).

It is sometimes permissible to express the relationship between R and t by the formula:

$$R = \frac{R_0}{1 - \dfrac{t}{\tau}} = \frac{R_0\tau}{\tau - t}.$$

This leads us to the equation:

$$\frac{di}{dt} + \frac{R_0\tau}{L(\tau - t)}t - \frac{v}{L} = 0.$$

To express t in parts of τ, we need to introduce a new variable x instead of t, according to the formula:

$$t = \tau x,$$

where x varies from $x = 0$ (initial instant) to $x = 1$ (the instant of quenching the spark, of breaking contact). The equation takes the form:

$$\frac{di}{dx} + \frac{R_0\tau}{L(1 - x)}i - \frac{v\tau}{L} = 0 \tag{39}$$

with the condition:

$$i|_{x=0} = i_0 \qquad \left(i_0 = \frac{v}{R_0}\right).$$

On applying (28), we easily obtain the general solution of the equation

$$i = (1 - x)^{\frac{R_0\tau}{L}}\left[\frac{v\tau}{L}\int(1 - x)^{-\frac{R_0\tau}{L}}dx + C\right],$$

where two cases can be distinguished:

$$1)\ \frac{L}{R_0} \neq \tau \qquad 2)\ \frac{L}{R^0} = \tau.$$

We find in case 1):

$$i = \frac{v\tau}{R_0\tau - L}(1 - x) + C(1 - x)^{\frac{R_0\tau}{L}}$$

and we determine the arbitrary constant C on substituting $x = 0$:

$$i_0 = \frac{v\tau}{R_0\tau - L} + C; \qquad C = i_0 - \frac{v\tau}{R_0\tau - L},$$

and finally,

$$i = \frac{v\tau}{R_0\tau - L}(1-x) + \left(i_0 - \frac{v\tau}{R_0\tau - L}\right)(1-x)^{\frac{R_0\tau}{L}}. \tag{40_1}$$

We proceed similarly and find in case 2):

$$i = (1-x)\left[i_0 - \frac{v\tau}{L}\log(1-x)\right]. \tag{40_2}$$

Bernoulli's equation is a generalisation of the linear differential equation (25):

$$y' + P(x)\,y + Q(x)\,y^m = 0, \tag{41}$$

where the exponent m can be considered as differing from zero and unity, since the equation is linear in these cases. We divide both sides by y^m:

$$y^{-m}y' + P(x)\,y^{1-m} + Q(x) = 0$$

and we introduce a new unknown function u instead of y:

$$u = y^{1-m}; \quad u' = (1-m)\,y^{-m}y'.$$

The equation now reduces to the form:

$$u' = P_1(x)\,u + Q_1(x) = 0\,,$$

where

$$P_1(x) = (1-m)\,P(x) \text{ and } Q_1(x) = (1-m)\,Q(x)\,,$$

i.e. *Bernoulli's equation reduces to a linear equation by substituting* $u = y^{1-m}$ and is then integrated as a linear equation.

We remark that integration of the differential equation of the form:

$$y' + P(x)\,y + Q(x)\,y^2 + R(x) = 0\,, \tag{41_1}$$

known as Riccati's equation, does not reduce to quadrature in the case of arbitrarily chosen coefficients. It can reduce to a linear equation if any one particular solution is known. Let $y_1(x)$ be in fact a solution of equation (41_1), i.e.:

$$y_1' + P(x)\,y_1 + Q(x)\,y_1^2 + R(x) = 0. \tag{*}$$

We introduce into (41_1) a new required function u instead of y, where

$$y = y_1 + \frac{1}{u}\,.$$

On substituting in (41_1) and taking into account equation (*), we obtain a linear equation for u of the form:

$$u' - \left[P(x) + 2Q(x)\,y_1\right]u - Q(x) = 0\,.$$

The general solution of this equation has the form: $u = C\varphi(x) + \psi(x)$. If we substitute this expression for u in the equation for y written above, we get the general solution of Riccati's equation in the form:

$$y = \frac{C\varphi_1(x) + \psi_1(x)}{C\varphi_2(x) + \psi_2(x)}.$$

5. Finding the solution of a differential equation with a given initial condition. As we have said, a first order differential equation

$$y' = f(x, y) \tag{42}$$

consists of a relationship between the coordinates (x, y) of a point and the slope y' of the tangent at this point. We assume that $f(x, y)$ is a single-valued, continuous function of (x, y). A definite tangent with a slope equal to $f(x, y)$ now corresponds, by (42), to any point of the plane at which $f(x, y)$ is defined. On indicating the direction of this tangent by an arrow passing through the corresponding point, we arrive at a *tangent field* in the plane, every tangent being associated with some point of the plane. The integral curves of equation (42) are the curves the tangents of which are the tangents of the field and they may be designated the *integral curves of the given field*.

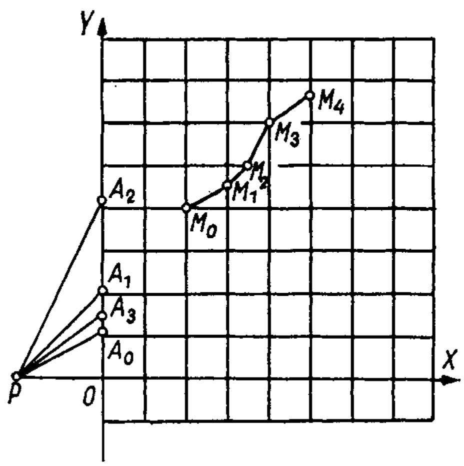

FIG. 8

The magnetic field at the earth's surface may be taken as an example. If we regard a portion of the earth's surface as a plane, the direction shown by the magnetic needle at each point gives us a definite tangent at every point.

We now turn to the question of finding the integral curves of equation (42). The complete definition of the position of an integral curve requires the further assigning of some point through which the integral curve must pass, e.g. its intersection with the line $x = x_0$ parallel to OY; or, what amounts to the same thing, we must assign the initial value y_0 that the required function y must take for the specified value $x = x_0$:

$$y\,|_{x=x_0} = y_0.$$

The integral curve passing through the given point (x_0, y_0) can be drawn approximately by using Euler's method, explained below.

We mark out a mesh of small equal squares in the coordinate plane by lines parallel to the axes, then we draw from the origin, in the negative direction of OX, the intercept $\overline{OP}$, of unit length (Fig. 8). We substitute $x = x_0$ and $y = y_0$ in the right-hand side of equation (42) and having found the value of $f(x_0, y_0)$, we mark off the intercept $\overline{OA}_0$ equal to this value on the ordinate axis. The line PA_0 will evidently have a slope equal to $f(x_0, y_0)$ and will therefore be parallel to the tangent to the integral curve at the point (x_0, y_0). We now proceed to the approximate construction of the integral curve itself in the form of a step line.

We produce from the point (x_0, y_0) a line $M_0 M_1$, parallel to PA_0 and hence having a slope $y_0' = f(x_0, y_0)$. Let $M_1(x_1, y_1)$ be the first point of intersection of this line with any side of our square mesh. We cut off a segment $\overline{OA}_1$ on the ordinate axis equal to $f(x_1, y_1)$, and produce through the point $M_1(x_1, y_1)$ a line $M_1 M_2$, parallel to PA_1 [and therefore having a slope $y_1' = f(x_1, y_1)$], to its first intersection at $M_2(x_2, y_2)$ with a side of our square mesh, and so on. This construction can be carried out both in the direction of increasing, and in the direction of decreasing, abscissae. The step line obtained in this way represents approximately the required integral curve.

We further remark that a different scale can be used for drawing the intercepts $\overline{OP}$ and $\overline{OA}_0, \overline{OA}_1, \ldots$ than is employed for the coordinates x and y, since the directions of $\overline{PA}_0, \overline{PA}_1, \ldots$ are evidently independent of the choice of scale for the intercepts.

This construction makes it clear by inspection that one and only one integral curve of equation (42) passes through a given point (x_0, y_0).

This assertion is susceptible of rigorous proof if the function $f(x, y)$ has properties in addition to continuity. For instance, *if $f(x, y)$ is a single-valued, continuous function of its arguments in the neighbourhood of the point (x_0, y_0) and has a continuous derivative with respect to y, one and only one integral curve of equation* (42) *passes through the point (x_0, y_0).*

This theorem, which at present we accept without proof, is usually called *the existence and uniqueness theorem for the solution of a differential equation with a given initial condition*. The theorem is proved at the end of the next chapter.

We supplement the above geometrical explanation with an analytic explanation of the theorem in an important particular case, viz, in the case where the right-hand side of equation (42) consists of a series expansion into positive integral powers of the differences $(x - x_0)$ and $(y - y_0)$ [I, 161]:

$$f(x, y) = \sum_{p, q=0}^{\infty} a_{pq} (x - x_0)^p (y - y_0)^q ,$$

which is convergent if the absolute values of the differences are sufficiently small.

Here, the solution of equation (42) satisfying the initial condition

$$y\,|_{x=x_0} = y_0 , \tag{43}$$

can be written as a Taylor series in positive integral powers of the difference $(x - x_0)$, the coefficients of the series being completely defined by equation (42). In fact, on substituting $x = x_0$ and $y = y_0$ in the right-hand side of (42), we obtain the value y_0 of the first derivative y' at $x = x_0$. We get on differentiating (42) with respect to x:

$$y'' = \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} y' ;$$

if we substitute $x = x_0$, $y = y_0$, $y' = y'_0$ in the right-hand side of this equation, we find the value y''_0 of the second derivative y'' at $x = x_0$. Further differentiation of the equation written above with respect to x gives us an equation in y''' and so on. We thus determine the Taylor series:

$$y = y_0 + \frac{y'_0}{1!}(x - x_0) + \frac{y''_0}{2!}(x - x_0)^2 + \dots , \tag{44}$$

which in fact gives, for values of x near x_0, the solution of (42) satisfying the initial condition (43).

The method of undetermined coefficients may be used as an alternative to the above method of determining successively the derivatives at $x = x_0$. We replace y on both sides of (42) by a power series with undetermined coefficients:

$$y = y_0 + a_1 (x - x_0) + a_2 (x - x_0)^2 + \dots \tag{45}$$

By expanding the right-hand side in powers of $(x - x_0)$ and equating coefficients of like powers of $(x - x_0)$, the coefficients $a_1, a_2, \dots$ can be successively determined. It can easily be shown that series (44) and (45) are identical.

Example. We find the solution of the equation:

$$y' = \frac{xy}{2}, \tag{46}$$

satisfying the initial condition:

$$y\,|_{x=0} = 1, \tag{47}$$

as a power series:

$$y = 1 + \sum_{s=1}^{\infty} a_s x^s,$$

where the constant term has been taken epual to unity in view of the initial condition (47).

We differentiate the series:

$$y' = \sum_{s=1}^{\infty} s a_s x^{s-1}.$$

We substitute these expressions for y and y' in equation (46):

$$a_1 + 2a_2 x + 3a_3 x^2 + \ldots + (n+1)\, a_{n+1} x^n + \ldots =$$

$$= \frac{1}{2} x\,(1 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1} + \ldots).$$

We equate coefficients of like powers of x on both sides and obtain the relationships shown in the table. It is clear from these that

x^0	$a_1 = 0$
x^1	$2a_2 = \frac{1}{2}$
x^2	$3a_3 = \frac{1}{2} a_1$
x^3	$4a_4 = \frac{1}{2} a_2$
. .	. . . .
x^n	$(n+1)\, a_{n+1} = \frac{1}{2} a_{n-1}$
. .	. . . .

$$a_1 = a_3 = a_5 = \ldots = a_{2n+1} = \ldots = 0;$$

$$a_2 = \frac{1}{4};\; a_4 = \frac{1}{2!\,4^2};\; \ldots;\; a_{2n} = \frac{1}{n!\,4^n};$$

i.e. finally [1, 126]

$$y = 1 + \frac{x^2}{4} + \frac{1}{2!}\left(\frac{x^2}{4}\right)^2 + \frac{1}{3!}\left(\frac{x^2}{4}\right)^3 +$$

$$+ \ldots + \frac{1}{n!}\left(\frac{x^2}{4}\right)^n + \ldots = e^{\frac{x^2}{4}}.$$

6. The Euler–Cauchy method. The approximate construction for an integral curve of equation (42) given in the previous article can be simplified by using lines only, parallel to OY, instead of the mesh of squares. This modified Eulerian method results in a relatively simple and handy means of evaluating approximately the ordinate y of an integral curve for a previously assigned abscissa x.

Let $M_0(x_0, y_0)$ be the initial point of the integral curve (Fig. 9). We produce a line with a slope $f(x_0, y_0)$ from this point to its intersection with the line $x = x_1$, parallel to OY, in the point M_1. Let y_1 be the ordinate of M_1. It is evidently given by the relationship:

$$y_1 - y_0 = f(x_0, y_0)(x_1 - x_0),$$

since $\overline{M_0N}$ and $\overline{NM_1}$ are given by the numbers (x_1-x_0) and (y_1-y_0), whilst the tangent of angle NM_0M_1 is equal to $f(x_0, y_0)$ by construction.

We draw $M_1 M_2$ with slope $f(x_1, y_1)$ from the point (x_1, y_1) to its intersection at M_2 with the next line $x = x_2$, parallel to OY. The

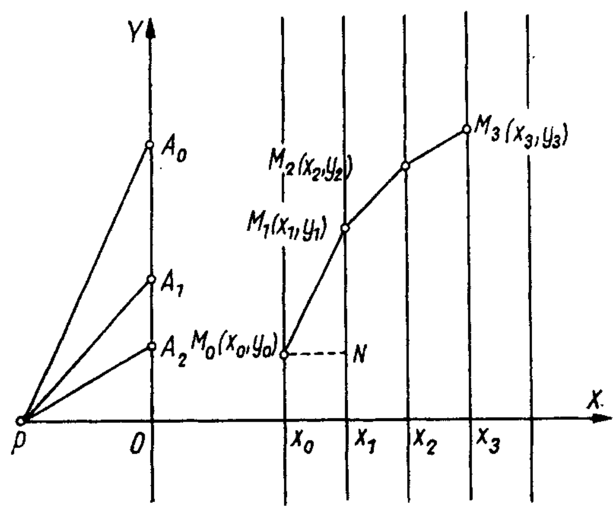

FIG. 9

ordinate of the point of intersection will be given by a relationship similar to the above:

$$y_2 - y_1 = f(x_1, y_1)(x_2 - x_1).$$

Proceeding in the same way from the point $M_2(x_2, y_2)$, we can next obtain the point $M_3(x_3, y_3)$ and so on. The lines PA_0, PA_1, ... have the same role in Fig. 9 as in Fig. 8.

We now suppose that, for a given value of x, we have to find the value y of the solution of equation (42) that satisfies the initial condition (43). By what was said above, we must proceed as follows: we subdivide the interval (x_0, x) by points:

$$x_0 < x_1 < x_2 < x_3 < \ldots < x_{n-2} < x_{n-1} < x \ldots \qquad (48)$$

and then obtain the ordinates $y_1, y_2, \ldots, y_{n-1}$ in accordance with the formulae:

$$\left.\begin{aligned} y_1 - y_0 &= f(x_0, y_0)(x_1 - x_0) \\ y_2 - y_1 &= f(x_1, y_1)(x_2 - x_1) \\ y_3 - y_2 &= f(x_2, y_2)(x_3 - x_2) \\ &\ldots\ldots\ldots\ldots\ldots \\ y_{n-1} - y_{n-2} &= f(x_{n-2}, y_{n-2})(x_{n-1} - x_{n-2}) \\ Y - y_{n-1} &= f(x_{n-1}, y_{n-1})(x - x_{n-1}). \end{aligned}\right\} \qquad (49)$$

With the conditions laid down in [5] for the properties of the function $f(x, y)$, if the number of sub-intervals increases, with each tending to zero, the quantity Y obtained from (49) will tend to the true ordinate y of the required integral curve provided the given x is sufficiently close to the initial x_0.

We easily find, on adding equations (49) term by term:

$$y \sim Y = y_0 + f(x_0, y_0)(x_1 - x_0) + f(x_1, y_1)(x_2 - x_1) + \ldots + \\ + f(x_{n-2}, y_{n-2})(x_{n-1} - x_{n-2}) + f(x_{n-1}, y_{n-1})(x - x_{n-1}). \qquad (50)$$

In the elementary case of the equation:

$$y' = f(x)$$

the formula written takes the form:

$$y_0 + \sum_{s=0}^{n-1} f(x_s)(x_{s+1} - x_s).$$

which, as we know [I, 87], gives an approximate expression for the value of the integral $y_0 + \int\limits_{x_0} f(x)\,dx$, i.e. for the solution of the given equation.

The computation in accordance with (49) is carried out in the following order. The first of equations (49) gives the difference $(y_1 - y_0)$. We add this to y_0, obtain the second ordinate y_1 then find the difference $(y_2 - y_1)$ with the aid of the second of equations (49). We add this latter to y_1, obtain the third ordinate y_2 then find $(y_3 - y_2)$ with the aid of the third of equations (49), and so on. We find Y by adding all these differences to y_0.

Example. We apply the method given to the solution of equation (46) with initial condition (47). We shall take all the intervals (x_0, x_1), (x_1, x_2) ... equal to 0.1.

x	y	$\frac{xy}{2}$	$\Delta y = \frac{xy}{2}\cdot 0.1$	$e^{\frac{x^2}{4}}$
0	1	1	0	1
0.1	1	0.05	0.005	1.0025
0.2	1.005	0.1005	0.0101	1.0100
0.3	1.0151	0.1523	0.0152	1.0227
0.4	1.0303	0.2061	0.0206	1.0408
0.5	1.0509	0.2627	0.0263	1.0645
0.6	1.0772	0.3232	0.0323	1.0942
0.7	1.1095	0.3883	0.0388	1.1303
0.8	1.1483	0.4593	0.0459	1.1735
0.9	1.1942	0.5374	0.0537	1.2244

The results of the computation are shown in the accompanying table. The first column contains x, the second contains the corresponding y, the third the value of $f(x, y)$, i.e. $xy/2$, the fourth the difference $\Delta y = y_{s+1} - y_s$, and the last the value of the ordinate of the accurate integral curve $y = e^{x^2/4}$.

As can be seen from the table, the error with $x = 0.9$ is less than 0.031, i.e. amounts to roughly 2.5%.

7. The general solution. On altering the value of y in the initial condition:

$$y|_{x=x_0} = y_0$$

we obtain an infinite set of solutions of equation (42), or in geometrical terms, a family of integral curves depending on the arbitrary constant y_0, this being the ordinate of the point of intersection of an integral curve with the line $x = x_0$. Instead of appearing in the solution as the initial value of y, the arbitrary constant can also appear in the general form:

$$y = \varphi(x, C). \tag{51}$$

Such a solution of (42), including the arbitrary constant, is called the general solution of the equation, as already mentioned [1]. It can also appear in implicit form:

$$\psi(x, y, C) = 0. \tag{52}$$

If we assign a definite numerical value to the constant C, we obtain a definite solution of (42), which is referred to as a *particular solution of the equation.* To distinguish the curve passing through a given point (x_0, y_0) from the family of curves of the general solution (52), we have to find the numerical value of C from the condition:

$$\psi(x_0, y_0, C) = 0. \tag{53}$$

The following is the converse of the problem of integrating a first order differential equation: *given the family of curves* (52), *depending on a single parameter C, it is required to form the differential equation for which this family is the family of the general integral.*

We get on differentiating the given equation (52) with respect to x:

$$\frac{\partial\psi(x, y, C)}{\partial x} + \frac{\partial\psi(x, y, C)}{\partial y} y' = 0. \tag{54}$$

Elimination of parameter C from equations (52) and (54) gives us the required differential equation of family (52):

$$\Phi(x, y, y') = 0.$$

After solution with respect to the arbitrary constant, the general solution (52) can be written in the form:

$$\omega(x, y) = C. \tag{55}$$

We obtain the general solution in this form in the case of the equation with separable variables [2]. The function $\omega(x, y)$ on the left-hand side of (55), is called a solution of the differential equation (42).

We must obtain a constant on substituting any particular solution of (42) for y in this function, i.e. *the solution of* (42) *is a function of x and y such that its total derivative with respect to x is zero, by virtue of* (42).

On taking the total derivative with respect to x of both sides of equation (55), we get [I, 69]:

$$\frac{\partial\omega(x, y)}{\partial x} + \frac{\partial\omega(x, y)}{\partial y} y' = 0,$$

or, on replacing y' by $f(x, y)$, inasmuch as y is a solution of (42) by hypothesis, we have:

$$\frac{\partial\omega(x, y)}{\partial x} + \frac{\partial\omega(x, y)}{\partial y} f(x, y) = 0. \tag{56}$$

The function $\omega(x, y)$ must satisfy this equation independently of the precise solution of (42) that we have substituted in this function. But in view of the arbitrariness of the initial condition (43) in the existence and uniqueness theorem, we can take any values we please of x and y, provided we take all the solutions of equation (42), i.e. *the function $\omega(x, y)$ must satisfy equation* (56) *as an identity in x and y.* We finally show how a solution of equation (42) can be checked when it is given implicity:

$$\omega_1(x, y) = 0. \tag{57_1}$$

We obtain as above the equation:

$$\frac{\partial \omega_1(x, y)}{\partial x} + \frac{\partial \omega_1(x, y)}{\partial y} f(x, y) = 0, \tag{57_2}$$

which must be satisfied at all points of curve (57_1), i.e. equation (57_2) is to be satisfied only by virtue of (57_1) and not as an identity in x and y: in short, (57_2) must be a consequence of (57_1).

Example.
We take, for instance, the equation:

$$y' = \frac{1 - 3x^2 - y^2}{2xy}.$$

It is easily shown that the circle:

$$x^2 + y^2 - 1 = 0$$

is a solution of this equation. Here, in fact, $f(x, y) = (1 - 3x^2 - y^2)/2xy$ and $\omega_1(x, y) = x^2 + y^2 - 1$, so that (57_2) reads:

$$2x + 2y\frac{1 - 3x^2 - y^2}{2xy} = 0, \quad \text{i.e.} \quad \frac{1 - x^2 - y^2}{x} = 0,$$

which is evidently satisfied by virtue of the equation of the circle. We show that the general integral of the given differential equation is:

$$x^3 + xy^2 - x = C.$$

We get by substituting in (56) $\omega_1(x, y) = x^3 + xy^2 - x$:

$$3x^2 + y^2 - 1 + 2xy\frac{1 - 3x^2 - y^2}{2xy} = 0,$$

and it is obvious that this equation is satisfied identically for all x and y.

Let the differential equation be given implicitly with respect to y':

$$\Phi(x, y, y') = 0. \tag{58}$$

If we solve it for y', we reduce it to form (42), though $f(x, y)$ can now be a many-valued function. We suppose that the function has m different values, so that there are m different values of y' for a given x and y, i.e. instead of a single tangent corresponding to a given point, we have m different tangents. As a result, we now have m different tangent fields in the plane instead of one tangent field. An integral curve passes through a given point for each of these fields, so that altogether m integral curves of equation (58) will pass through the given point. Yet the general integral of (58) will contain only one arbitrary constant, i.e. will have the form (52); on the other hand, equation (53) must in general give m distinct values, and not one value, for C.

We make up an example in connection with these last remarks, where the solution containing an arbitrary constant is not strictly speaking the general solution. We take the differential equation:

$$y'^2 - xy' = 0. \tag{59}$$

The left-hand side can be factorized, giving $y'(y' - x) = 0$, so that in essence we have two distinct differential equations:

$$y' = 0 \text{ and } y' - x = 0,$$

with general solutions

$$y - C = 0 \tag{59$_1$}$$

and

$$y - \frac{1}{2}x^2 - C = 0. \tag{59$_2$}$$

The last two equations can be combined:

$$(y - C)\left(y - \frac{1}{2}x^2 - C\right) = 0,$$

giving the general solution of equation (59). Two integral curves pass through every point of the plane: the straight line (59_1) and the parabola (59_2). Evidently (59_1), $y = C$, gives a solution of (59) containing an arbitrary constant; this solution is not the general solution of (59), but only the general solution of the equation $y' = 0$.

Equation (42), or (58), can have a solution which is not contained in the family of the general solution, i.e. cannot be obtained from (52) with some particular value of constant C. Such a solution is called a singular solution of the equation. We go into the problem of finding such solutions, and their geometrical interpretation, in [10].

Strictly speaking, the concepts of solution and general solution are in need of further explanation. We shall not go into the matter,

however, inasmuch as the existence and uniqueness theorem for the solution with a given initial condition is a more natural basis for a theoretical treatment of differential equations. Finding the general solution, as described above for a particular type of equation, certainly offers a very useful practical means of constructing the solutions of differential equations. We remark here that if, on passing from the differential equation to its general solution, we at no step violate the equivalence of succesive equations, there can be no singular solutions, i.e. every solution is contained in the general solution, on assigning various numerical values to C. In the case when the equivalence of the equations is lost, the singular solutions must be sought among the missing solutions, as will be done in [8] and [9].

By general solution is naturally understood a solution of the differential equation containing an arbitrary constant, from which can be obtained all the solutions defined by the existence and uniqueness theorem for initial conditions filling a certain domain of the (x, y) plane. This domain is determined by the function $f(x, y)$ appearing in equation (42). It is natural to describe solutions of the differential equation as singular solutions when they have the property that the conditions guaranteeing the existence and uniqueness theorem are not fulfilled at any point of the corresponding integral curve. All these definitions require certain assumptions, of course, regarding the function $f(x, y)$ or $\Phi(x, y, y')$ appearing in equation (42) or (58).

On replacing y' by the arbitrary constant C_1 in equation (42) or (58), we get the family of curves:

$$f(x, y) = C_1 \quad \text{or} \quad \Phi(x, y, C_1) = 0.$$

Each curve of this family is the locus of points of the plane which are associated with the same slope, the family as a whole being referred to as a family of isoclines of the given differential equation, i.e. a family of curves of the same slope. In the particular case of the magnetic field at the earth's surface, the isoclines are lines along which the direction of the magnetic needle is constant.

The isoclines for the homogeneous equation of [3] were lines passing through the origin.

We shall note the cases in which an isocline is an integral curve of the equation, i.e. gives a solution of the equation. We take the isocline:

$$\Phi(x, y, b) = 0.$$

corresponding to the particular value $C_1 = b$. At points of the isocline, the differential equation gives the same slope, inasmuch as $y' = b$. A necessary and sufficient condition for the isocline to be a solution is that the tangent to the isocline is also of slope b at every point of it — whence it immediately follows that the isocline must be a straight line of slope b, since $y' = b$ gives $y = bx + c$, where c is a constant. Hence, *an isocline is a solution only when it is a straight line and when the direction of this line coincides with the constant direction of the tangents, as defined by the differential equation at points of the isocline.*

Example. To find the curves for which the length of the normal MN is a constant a (Fig. 10). Use of the expression for the length of the normal [I, 77] gives us the differential equation:

$$\pm\, y\sqrt{1 + y'^2} = a. \tag{60}$$

We get by squaring both sides of the equation and solving with respect to y':

$$\frac{dy}{dx} = \pm \frac{\sqrt{a^2 - y^2}}{y}. \tag{61}$$

The right-hand side of the last equation is only defined for $|y| < a$, i.e. in the strip between the lines

$$y = a \text{ and } y = -a, \tag{62}$$

since otherwise the expression under the square root is negative; at every point inside the strip, y' has two distinct values.

The variables are separable in equation (61):

$$\frac{y\,dy}{\sqrt{a^2 - y^2}} = \pm\, dx. \tag{63}$$

We easily find on integrating:

$$(x - C)^2 + y^2 = a^2, \tag{64}$$

i.e. the family of circles with centres on OX and radius equal to a (Fig. 10). All these circles are situated in the strip bounded by the straight lines (62), with two circles of family (64) passing through every point inside the strip.

The transition from equation (61) to (63) required division by $\sqrt{a^2 - y^2}$, and as a result of this the solution $y = \pm a$ might have been lost. It is easily seen by direct substitution that this is in fact a solution of (61). The solution is represented geometrically by the lines (62), which are not included in the family of the general solution (64); in other words, the solution cannot be found from (64) whatever the value of constant C, i.e. it is a singular solution.

Substitution of the constant C_1 for y' in equation (60) gives us the family of isoclines:

$$\pm y\sqrt{1+C_1^2}=a,$$

These are lines parallel to OX. The tangents to circles (64) along these lines maintain a constant direction.

The lines (62), in particular, are also isoclines, with y' maintaining a constant value zero along them, which coincides with the slope of the lines themselves: so that the lines are at the same time solutions of equation (61).

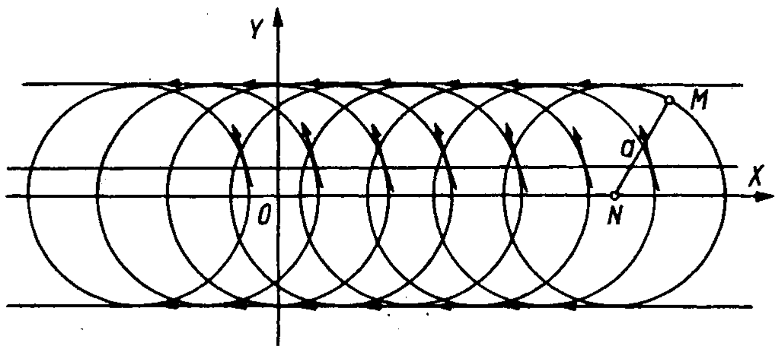

FIG. 10

Inside the strip given by the lines (62) we have two differential equations (61): one corresponding to the (+) sign and the other to the (−) sign. The circles (64) inside the strip are obtained in accordance with the existence and uniqueness theorem. The theorem becomes inapplicable at points of the lines $y = \pm a$, and these represent singular solutions of equation (60) or (61).

8. Clairaut's equation. An equation of the type

$$y = xy' + \varphi(y') \tag{65}$$

is called a Clairaut equation. Substitution of an arbitrary constant C for y' gives us a family of isoclines of the equation:

$$y = xC + \varphi(C). \tag{66}$$

Every isocline is seen to be a straight line, with slope equal to the constant that we substituted for y', i.e. the direction of each of lines (66) is the same as the constant tangential direction defined by the differential equation at points of the line. Recalling what was said in the previous article, we can assert that each of lines (66) is also a solution of equation (65), i.e. the family of isoclines (66) is at the same time the family of the general solution of (65).

We now indicate a second method of obtaining the general solution of equation (65), whereby the singular solution of the equation is

found, as well as its general solution. We use the notation $y' = p$, and re-write (65):

$$y = xp + \varphi(p). \tag{67}$$

It amounts to finding p as a function of x, say $\omega(x)$, so that on substituting $p = \omega(x)$ on the right-hand side of (67) we get for y a function of x such that its derivative y' is: $y' = p = \omega(x)$. We take differentials of both sides of (67), expand the left-hand side as $\mathrm{d}y = y'\mathrm{d}x = p\mathrm{d}x$, and obtain the first order differential equation for p:

$$p\mathrm{d}x = p\mathrm{d}x + x\mathrm{d}p + \varphi'(p)\mathrm{d}p \quad \text{or} \quad [x + \varphi'(p)]\,\mathrm{d}p = 0.$$

We get two cases on equating each factor to zero. The case $\mathrm{d}p = 0$ gives $p = C$, where C is an arbitrary constant; substitution of $p = C$ in equation (67) again gives us the general solution (66). In the second case we have the equation:

$$x + \varphi'(p) = 0. \tag{68}$$

On eliminating p from (67) and (68), i.e. from the two equations:

$$y = xp + \varphi(p) \quad \text{and} \quad x + \varphi'(p) = 0, \tag{69}$$

we likewise obtain a solution of equation (65), which does not contain an arbitrary constant. This is usually a singular solution of the equation.

The geometrical problem of finding the curve, given the properties of its tangent, reduces to Clairaut's equation, assuming that the properties relate only to the tangent itself, and not to the point of contact. The equation of the tangent has the form:

$$Y - y = y'(X - x) \quad \text{or} \quad Y = y'X + (y - xy'),$$

and any properties of the tangent are expressed by a relationship between $(y - xy')$ and y':

$$\Phi(y - xy',\ y') = 0.$$

On solving with respect to $(y - xy')$, we arrive at an equation of the form (65). The straight lines composing the general solution of Clairaut's equation are evidently of no interest as regards providing an answer to our geometrical problem, the answer being in fact given by the singular solution of the equation.

Example. To find the curve such that the intercept T_1T_2 cut off its tangent by the coordinate axes is of constant length a (Fig. 11).

The equation of the tangent gives us the projections OT_1 and OT_2 of the tangent on the coordinate axes, and this enables us to write the differential equation of the required curve as:

$$\frac{(y - xy')^2}{y'^2} + (y - xy')^2 = a^2 \quad \text{or} \quad y = xy' \pm \frac{ay'}{\sqrt{1 + y'^2}}\,.$$

The general solution is:

$$y = xC \pm \frac{aC}{\sqrt{1 + C^2}}\,. \tag{70}$$

consisting of a family of straight lines, the length of the intercepts of which on the axes is equal to a. The singular solution is obtained as a result of eliminating p from:

$$y = xq \pm \frac{ap}{\sqrt{1 + C^2}} \tag{71}$$

and from the equation

$$x \pm \frac{\sqrt{1 + p^2} - \dfrac{p^2}{\sqrt{1 + p^2}}}{1 + p^2} = 0,$$

which reduces to:

$$x \pm \frac{a}{(1 + p^2)^{3/2}} = 0.$$

We write $p = \tan \varphi$, giving

$$x = \mp a \cos^3 \varphi$$

whilst equation (71) for y gives us:

$$y = \mp a \cos^3 \varphi \tan \varphi \pm a \sin \varphi = \pm a \sin^3 \varphi$$

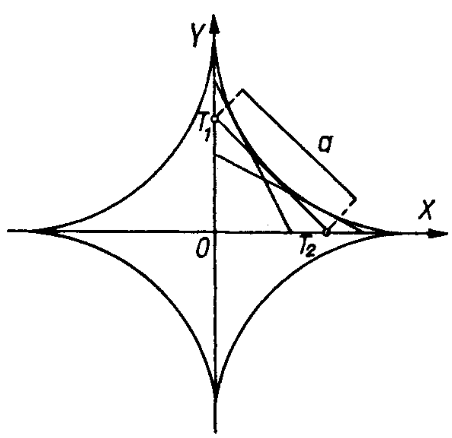

FIG. 11

We eliminate φ by raising the last two equations to the power 2/3 and adding:

$$x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}},$$

i.e. the required curve is an astroid, which we mentioned in [1, 80]. The straight lines (70) form the family of tangents to it (Fig. 11).

9. Lagrangian equations. An equation of the form:

$$y = x\varphi_1(y') + \varphi_2(y'). \tag{72}$$

is called a Lagrangian equation, $\varphi_1(y')$ being assumed different from y'; if $\varphi_1(y') = y'$, we get the Clairaut equation just described.

We use the same method of differentiation for (72) as for the Clairaut equation. We write $y' = p$, so that the equation becomes

$$y = x\varphi_1(p) + \varphi_2(p). \tag{73}$$

We take the differentials of both sides and get a first order equation for p:

$$p\mathrm{d}x = \varphi_1(p)\,\mathrm{d}x + x\varphi_1'(p)\,\mathrm{d}p + \varphi_2'(p)\,\mathrm{d}p.$$

Division by $\mathrm{d}p$ gives us the equation:

$$[\varphi_1(p) - p]\frac{\mathrm{d}x}{\mathrm{d}p} + \varphi_1'(p)\,x + \varphi_2'(p) = 0,$$

which, on the assumption that x is a function of p, is a linear differential equation. We reduce this to the form (25) by dividing both sides by $[\varphi_1(p) - p]$, and obtain its general solution in the form:

$$x = \psi_1(p)\,C + \psi_2(p). \tag{74}$$

Substitution of this expression for x in equation (72) gives us an equation for y of the form:

$$y = \psi_3(p)\,C + \psi_4(p). \tag{75}$$

Equations (74) and (75) express x and y in terms of an arbitrary constant C and a variable parameter p, i.e. give the general solution of the Lagrangian equation in parametric form. On eliminating parameter p from (74) and (75), we get the ordinary equation for the general solution.

When dividing the equation by $\mathrm{d}p$, we may have lost the solution corresponding to $\mathrm{d}p = 0$, i.e. corresponding to constant p, or what amounts to the same thing, to constant y'. But constant y' leads to a first degree polynomial for y, i.e. the missing solutions must be straight lines, if they exist. We also note that, for constant $p = a$, (73_1) gives $a\,\mathrm{d}x = \varphi_1(a)\,\mathrm{d}x$, i.e. the value of the constant a must be defined by the equation $\varphi_1(a) - a = 0$.

We give the geometrical interpretation of this last fact. Substitution of constant C_1 for y' in equation (72) gives us the equation of the isoclines:

$$y = x\varphi_1(C_1) + \varphi_2(C_1), \tag{76}$$

i.e. *the isoclines of a Lagrangian equation are straight lines*. The solutions which are represented by straight lines have to be sought among the isoclines. For this, we have to establish the condition that the slope $\varphi_1(C_1)$ of the isocline is the same as the constant slope C_1 of the tangent along the isocline:

$$\varphi_1(C_1) - C_1 = 0.$$

On solving this equation and substituting the value found for C_1 in equation (76), we obtain the required solutions, among which must be included the singular solution in question.

10. The envelope of a family of curves, and singular solutions. We have already had two examples in which singular solutions were obtained in addition to the general solution. The general solution in the example of [7] consisted of the family of circles

$$(x - C)^2 + y^2 = a^2 \tag{77}$$

with centres on OX and of fixed radius a.

The two lines $y = \pm a$, parallel to OX, were singular solutions. Any given point of these lines is a point of contact with a circle of family (77) (Fig. 10). The general solution in the example of [8] consisted of a family of straight lines whose intercepts cut off by the coordinate axes were equal in length to the given a, whilst the singular solution was the astroid, such that any given point of it was a point of contact with one of the lines concerned, i.e. the family of straight lines was a family of tangents to the astroid.

These examples lead us naturally to the concept of the envelope of a family of curves. Let the family of curves

$$\varphi(x, y, C) = 0, \tag{78}$$

be given, where C is an arbitrary constant. *The envelope of the family is defined as the curve, every point of which is a point of contact with a curve of the family,* i.e. *the tangent at any given point of the envelope is also a tangent to the curve of family* (78) *that passes through this point.*

We derive the rule for finding the envelope. We start by finding the slope of the tangent to a curve of family (78). We differentiate equation (78), whilst taking into account that y is a function of x and C is a constant; this gives us

$$\frac{\partial\varphi(x, y, C)}{\partial x} + \frac{\partial\varphi(x, y, C)}{\partial y}\,\frac{dy}{dx} = 0,$$

whence [I, 69]:

$$\frac{dy}{dx} = -\frac{\dfrac{\partial\varphi(x, y, C)}{\partial x}}{\dfrac{\partial\varphi(x, y, C)}{\partial y}}\,. \tag{79}$$

We assume that the required equation of the envelope is

$$R(x,y) = 0. \tag{80}$$

We can suppose that the left-hand side of this equation, $R(x, y)$, which is as yet unknown, has the form $\psi(x, y, C)$, where C, instead of being a constant, is some unknown function of x and y. For any given function $R(x, y)$, in fact, we can write the equality

$$R(x,y) = \psi(x,y,C),$$

which defines C for us as a function of x and y. In other words, we can look for the equation of the envelope in the form (78), except for C being a required function of x and y instead of being a constant. We differentiate both sides of (78), and obtain, since C is no longer constant:

$$\frac{\partial\psi(x,y,C)}{\partial x}\,\mathrm{d}x + \frac{\partial\psi(x,y,C)}{\partial y}\,\mathrm{d}y + \frac{\partial\psi(x,y,C)}{\partial C}\,\mathrm{d}C = 0. \tag{81}$$

The slope $\mathrm{d}y/\mathrm{d}x$ of the tangent to the envelope must, by hypothesis, be the same as that of the tangent to the curve of family (78) that passes through the same point, i.e. equation (81) must give us equation (79) above for $\mathrm{d}y/\mathrm{d}x$; but this can only be the case when the third term on the left-hand side of (81) vanishes, i.e. when $(\partial\psi(x,y,C)/\partial C)\,\mathrm{d}C = 0$. The possibility $\mathrm{d}C = 0$ gives us constant C, i.e. a curve of the family and not the envelope; so that to obtain the envelope we must put

$$\frac{\partial\psi(x,y,C)}{\partial C} = 0.$$

This equation also defines C as a function of (x, y). Substitution of the expression obtained for C in terms of x and y in the left-hand side of (78) gives us the equation (80) of the envelope, i.e. *the equation of the envelope of family* (78) *can be obtained by eliminating C from the two equations*:

$$\psi(x,\ y,\ C) = 0; \quad \frac{\partial\psi(x,\ y,\ C)}{\partial C} = 0. \tag{82}$$

As we move along the envelope, we touch different curves of family (78), each curve being defined by its value of constant C; this makes it clear why the equation of the envelope was sought in the form (78), with C, however, taken as variable.

We now turn to the singular solution of a differential equation. We let (78) be the family of the general solution of the differential equation:

$$\Phi(x,\ y,\ y') = 0, \tag{83}$$

i.e. the coordinates (x, y) and slope y' of the tangent for any given curve of family (78) satisfy equation (83). At every point of the envelope x, y and y' will coincide with the x, y and y' for some curve of family (78), i.e. the x, y and y' of the envelope will also satisfy (83). In other words, *the envelope of the family of the general solution is also an integral curve of the equation.*

If $\psi(x, y, C) = 0$ is the general solution of equation (83), elimination of C from equations (82) leads us to a singular solution in certain cases. We add the proviso here, "in certain cases" (and not always), due to the following considerations. It was assumed in the above arguments that curves (78) have tangents; therefore, if we eliminate C from equations (82), it is possible for us to obtain not only the envelope, but also the set of all the singular points of the curves of family (78), at which the curves do not possess definite tangents [I, 76]. Furthermore, it sometimes happens that the envelope itself enters into the constitution of family (78). We shall not give a rigorous treatment of the theory of envelopes and singular solutions. The theory must obviously be closely connected with the existence and uniqueness theorem, mentioned in [5]. We confine ourselves to explaining the problem in a few examples.

1. We seek the envelope of the family of circles (77):

$$(x - C)^2 + y^2 = a^2.$$

Equations (82) here take the form:

$$(x - C)^2 + y^2 = a^2; \quad -2(x - C) = 0.$$

The second equation gives $C = x$, and substituting this in the first equation gives us $y^2 = a^2$, i.e. the set of two straight lines $y = \pm a$, which we obtained previously.

2. The general solution of Clairaut's equation $y = xy' + \varphi(y')$ is

$$y = xC + \varphi(C).$$

The envelope is obtained by eliminating C from the two equations:

$$y = xC + \varphi(C); \quad 0 = x + \varphi'(C).$$

These equations coincide with equations (69) of [8], with the trivial replacement of the letter p by C, i.e. we get the previous rule for finding the singular solution of Clairaut's equation.

3. The curve $y^2 = x^3$ is the so-called semicubical parabola (Fig. 12). On displacing the curve parallel to OY, we get a family of semicubical parabolas:

$$(y + C)^2 = x^3.$$

Each of these curves has a cusp on OY, and there exists a right-hand tangent at the cusp, parallel to OX. Equations (82) here take the form:

$$(y + C)^2 = x^3; \quad 2(y + C) = 0.$$

Elimination of C gives us $x = 0$, i.e. axis OY. Axis OY is not the envelope in this case, but the locus of singular points of curves of the family.

4. We consider the family of curves

$$y = C(x - C)^2.$$

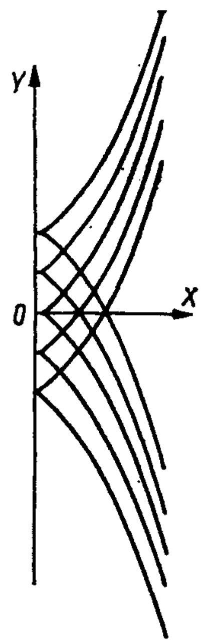

FIG. 12

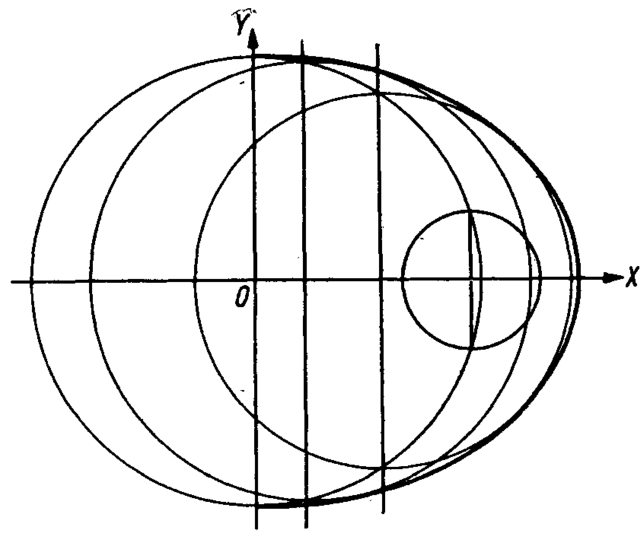

FIG. 13

We have a parabola for $C \neq 0$, and the axis OX for $C = 0$. Equations (82) become:

$$y = C(x - C)^2; \quad (x - C)(x - 3C) = 0.$$

The second equation gives $C = x$ or $C = x/3$. Substitution in the first equation gives us either $y = 0$ or $y = 4x^3/27$. The first curve $y = 0$ is axis OX, which belongs to the given family of curves; whereas the cubical parabola $y = 4x^3/27$ is the envelope of the family.

5. We take the chords of the circle of unit radius, centre at the origin, that are perpendicular to OX and we draw fresh circles with the chords as diameters, thus obtaining a family of circles. If $x = C$ is the point of intersection of a chord with OX, the square of the radius of the corresponding circle is $(1 - C^2)$ (Fig. 13), so that the equation of the family is:

$$(x - C)^2 + y^2 = 1 - C^2.$$

Differentiation with respect to C gives us the equation:

$$-2(x - C) = -2C;$$

on eliminating C from the last two equations, we get the equation:

$$\frac{x^2}{2} + y^2 = 1\,,$$

i.e. we obtain an ellipse with semi-axes $\sqrt{2}$ and 1, with the coordinate axes as axes of symmetry. It is obvious from the figure that this ellipse does not touch all the circles of the family.

11. Equations quadratic in y'. We consider in more detail, from the point of view of singular solutions, differential equations that are quadratic in y':

$$\Phi(x, y, y') = y'^2 + 2P(x, y)\,y' + Q(x, y) = 0\,, \tag{84}$$

where $P(x, y)$ and $Q(x, y)$ are single-valued and continuous, and have continuous derivatives with respect to y, throughout the domain; e.g. they may be polynomials in x and y. We obtain on solving with respect to y':

$$y' = -P(x, y) \pm \sqrt{R(x, y)}, \tag{85}$$

where we have taken $R(x, y) = [P(x, y)]^2 - Q(x, y)$. In the part of the domain where $R(x, y) > 0$, (85) is equivalent to two differential equations, and in accordance with the existence and uniqueness theorem, two and only two integral curves will pass through every point of this part of the domain. Differential equation (84) will have no singular solutions in this region. In the region where $R(x, y) < 0$, equation (85) does not yield a real y', and there are no integral curves in this region. Finally, we consider the equation

$$R(x, y) = 0\,, \tag{86}$$

which can define one or more curves in the domain. It is only among these curves that singular solutions of equation (84) can be found. We remark that (86) can be obtained by eliminating y' from (84) and the equation:

$$\frac{\partial\Phi(x, y, y')}{\partial y'} = 0\,. \quad \text{i.e.} \quad y' + P(x, y) = 0\,.$$

The latter equation expresses the fact that (84) has a multiple root with respect to y'.

1. In the case of the equation

$$y = xy' + y'^2, \quad \text{i.e.} \quad y'^2 + xy' - y = 0$$

(86) takes the form $x^2/4 + y = 0$, and the parabola $y = -x^2/4$ is a singular solution of the Clairaut equation written.

2. In the case of the equation

$$y'^2 + 2xy' + y = 0$$

(86) gives $y = x^2$. This parabola does not satisfy the equation written, so that the latter has no singular solution whatever.

12. Isogonal trajectories. An *isogonal trajectory* is defined as the family of curves intersecting the curves of the family

$$\psi(x, y, C) = 0 \tag{87}$$

at a given angle.

If the given angle is a right angle, the trajectory is called the *orthogonal trajectory*. We show that finding an isogonal trajectory leads to integrating a first order differential equation.

On eliminating C from the equations:

$$\psi(x,\ y,\ C) = 0; \qquad \frac{\partial\psi(x,\ y,\ C)}{\partial x} + \frac{\partial\psi(x,\ y,\ C)}{\partial y}\, y' = 0\,,$$

we obtain the differential equation of the given family (87) as in [7]:

$$\Phi(x,\ y,\ y') = 0. \tag{88}$$

We start by finding the orthogonal trajectory. In this case, the tangents to the required curves are perpendicular to the tangents to the curves of family (87) at the points of intersection of the curves, i.e. the slopes of the tangents to the trajectory are the reciprocals, with reversed sign, of the slopes of the tangents to the given family. Hence it follows that, *to obtain the differential equation of the orthogonal trajectory, we must replace y' by $(-1/y')$ in the differential equation of the given family.*

Finding the orthogonal trajectory thus reduces to integration of the equation:

$$\Phi\left(x,\ y_1,\ -\frac{1}{y_1'}\right) = 0\,,$$

where y_1 is the required function of x.

We now turn to the general problem of isogonal trajectories. Let φ be the constant angle at which the curves of the trajectory intersect the curves of family (87). Let y_1 denote, as before, the ordinate of the required curve; on using the formula for the tangent of the difference of two angles:

$$\tan\varphi = \tan(\psi_1 - \psi) = \frac{\tan\psi_1 - \tan\psi_1}{1 + \tan\psi\tan\psi_1}\,,$$

where $\tan\psi = y'$ is the slope of the tangent to a curve of (87) and $\tan\psi_1 = y'_1$ is the slope of the tangent to the required curve, we can write

$$\frac{y_1' - y'}{1 + y' y_1'} = \tan\varphi, \tag{89}$$

where φ is measured from curve (87) to the required curve. On eliminating y' from the last equation and equation (88), we obtain the differential equation of the isogonal trajectory, which then has to be integrated.

We come across orthogonal trajectories when considering *plane fluid flow*. We assume that the fluid flow takes place in a plane, so that a vector $\mathbf{v}$, the velocity of motion, is defined at every point (x, y) of the plane. If the velocity vector depends only on the position of the point in the plane, and not on time, the motion is described as *steady* or *established*. We shall confine ourselves to this type of motion. We further assume that there exists a velocity potential, i.e. that the projections of vector $\mathbf{v}(x, y)$ on the coordinate axes are the partial derivatives $\partial u(x, y)/\partial x$ and $\partial u(x, y)/\partial y$ of some function $u(x, y)$. The curves of the family

$$u(x, y) = C \tag{90}$$

are described in this case as *equipotential lines*.

The lines, the tangents to which have, at every point, the same direction as the vector $\mathbf{v}(x, y)$, are called *stream lines* and give the trajectories of the moving particles. We show that the stream lines form the *orthogonal trajectories of the family of equipotential lines*.

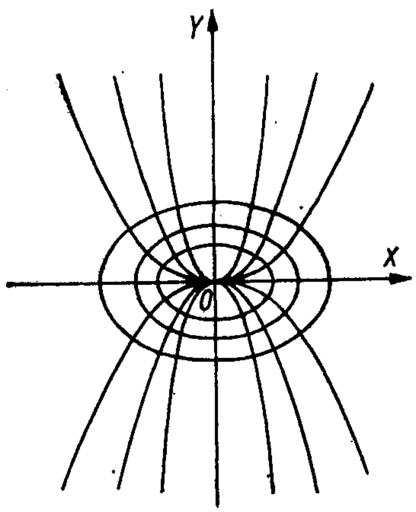

FIG. 14

Let φ be the angle formed by the velocity vector $\mathbf{v}(x, y)$ with axis OX, where $|\mathbf{v}|$ is the length of this vector. By hypothesis, $\partial u(x,y)/\partial x$ and $\partial u(x, y)/\partial y$ are the projections of $\mathbf{v}(x, y)$ on the axes, i.e.

$$\frac{\partial u(x, y)}{\partial x} = |\mathbf{v}| \cdot \cos\varphi \text{ and}$$

$$\frac{\partial u(x, y)}{\partial y} = |\mathbf{v}| \cdot \sin\varphi ,$$

whence we obtain the expression for the slope of the tangent to a stream line as:

$$\tan\varphi = \frac{\dfrac{\partial u(x, y)}{\partial y}}{\dfrac{\partial u(x, y)}{\partial x}}. \tag{91}$$

The slope of the tangent to an equipotential line (90) is found by differentiating this equation with respect to x:

$$\frac{\partial u(x, y)}{\partial x} + \frac{\partial u(x, y)}{\partial y} y' = 0 , \text{ whence } y' = -\frac{\dfrac{\partial u(x, y)}{\partial x}}{\dfrac{\partial u(x, y)}{\partial y}} ,$$

i.e. we obtain a slope which is the reciprocal, with reversed sign, of slope (91). Hence it follows that *the equipotential lines and the stream lines are orthogonal to each other*.

If a family of curves is a family of equipotential lines, its orthogonal trajectories form the family of corresponding stream lines, and vice versa. In the case of a plane electrostatic field, the lines of force represent the orthogonal trajectories of the family of equipotential lines.

Example. To find the isogonal trajectories of the family

$$y = Cx^m. \tag{92}$$

On eliminating C from the equations

$$y = Cx^m; \quad y' = Cmx^{m-1},$$

we get the differential equation of family (92):

$$y' = m\frac{y}{x}.$$

On substituting this expression for y' in (89), we get the differential equation of the required family:

$$\frac{y' - m\dfrac{y}{x}}{1 + m\dfrac{yy'}{x}} = \frac{1}{k},$$

the constant $\tan\varphi$ being written as $1/k$, and writing simply y instead of y_1. This equation reduces to the form:

$$y' = \frac{km\dfrac{y}{x} + 1}{k - m\dfrac{y}{x}} \tag{93}$$

and is therefore a homogeneous equation [3].

If $m = 1$, (92) is a family of radius vectors passing through the origin, and the required curves must cut these at a constant angle, i.e. they are either logarithmic spirals [I, 83] or circles.

If $m = -1$ and $k = 0$, the problem reduces to finding the orthogonal trajectories of the rectangular hyperbolas

$$xy = C. \tag{94}$$

Here, (93) reduces to the equation with separable variables:

$$\frac{dy}{dx} = \frac{x}{y} \quad \text{or} \quad x\,dx - y\,dy = 0.$$

Integration again gives a family of rectangular hyperbolas, referred in this case to the axes of symmetry:

$$x^2 - y^2 = C.$$

As may easily be seen, this family is obtained from the given family (94) by turning it through 45° about the origin. In general, for $k = 0$, (93) reduces to the form:

$$\frac{dy}{dx} = -\frac{x}{my}.$$

and its general solution is:

$$my^2 + x^2 = C\,,$$

i.e. the orthogonal trajectories of family (92) consist, for $m > 0$, of a family of similar ellipses, and for $m < 0$, of a family of similar hyperbolas. The orthogonal trajectories of the parabolas $y = Cx^2$ are illustrated in Fig. 14.

§ 2. Differential equations of higher orders; systems of equations

13. General principles. An ordinary differential equation of the nth order has the form:

$$\Phi(x,\ y,\ y',\ y'',\ \ldots,\ y^{(n)}) = 0\,, \tag{1}$$

or, on solving with respect to $y^{(n)}$:

$$y^{(n)} = f(x,\ y,\ y',\ y'',\ \ldots,\ y^{(n-1)}). \tag{2}$$

Every function y of the independent variable x that satisfies equation (1) or (2) is called a *solution of the equation*, whilst the actual task of finding the solutions of the equation is described as the *task of integrating the equation*. We take as an example the linear motion of a point-mass of mass m under the action of a force $\mathbf{F}$, which depends on time t, on the position of the point and on its velocity. If we take as axis OX the straight line along which the point moves, the force $\mathbf{F}$ can be considered as a given function of t, x and $\mathrm{d}x/\mathrm{d}t$. By Newton's law, the product of the mass of the particle and its acceleration must be equal to the force acting. This gives us the differential equation of motion:

$$m\,\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = \mathbf{F}\left(t,\ x,\frac{\mathrm{d}x}{\mathrm{d}t}\right). \tag{3}$$

Integration of this second order equation determines the relationship between x and t, i.e. the motion of the particle under the action of the given force. In order to obtain a definite solution of the problem, we must also specify the *initial conditions of the motion*, i.e. the position of the particle and its velocity at some initial instant, say at $t = 0$:

$$x\Big|_{t=0} = x_0; \quad \frac{\mathrm{d}x}{\mathrm{d}t}\Big|_{t=0} = x_0', \tag{4}$$

In the case of the nth order equation (1) or (2), the initial conditions consist in a specification of the function y and of its derivatives up to and including the $(n-1)$th order for a given value of $x = x_0$:

$$y|_{x=x_0} = y_0; \quad y'|_{x=x_0} = y'_0; \ldots; \quad y^{(n-1)}|_{x=x_0} = y_0^{(n-1)}. \tag{5}$$

The $y_0, y'_0, \ldots, y_0^{(n-1)}$ here are definitely assigned numbers.

A uniqueness and existence theorem is valid for the nth order equation, as for the first order equation, and can be stated as follows: *if $f(x, y, y', \ldots, y^{(n-1)})$ is a single-valued function of its arguments, is continuous for all x in the neighbourhood of x_0 and for all $y, y', \ldots, y^{(n-1)}$ in the neighbourhood of* (5), *and has continuous first order partial derivatives with respect to $y, y', \ldots, y^{(n-1)}$, a single definite solution of equation* (2) *corresponds to initial conditions* (5).

On varying the constants $y_0, y'_0, \ldots, y_0^{(n-1)}$ in the initial conditions, we obtain an infinite set of solutions, or more accurately, a family of solutions, depending on n arbitrary constants. These arbitrary constants can appear in the solution, not as initial conditions, but in the more general form:

$$y = \varphi(x, C_1, C_2, \ldots, C_n). \tag{6}$$

Such a solution of equation (2), *containing n arbitrary constants, is called the general solution of* (2). The equation of the general *solution* can also be written in implicit form:

$$\psi(x, y, C_1, C_2, \ldots, C_n) = 0. \tag{7}$$

On assigning definite values to constants $C_1, C_2, \ldots, C_n$, we obtain *particular solutions* of the equation.

We obtain n equations by differentiating equation (6) or (7) $(n-1)$ times with respect to x then substituting $x = x_0$ and initial conditions (5). It is assumed that these equations are soluble with respect to $C_1, C_2, \ldots, C_n$ for any given initial conditions $(x_0, y_0, y'_0, \ldots, y_0^{(n-1)})$ from a certain interval of variation of $x_0, y_0, y'_0, \ldots, y_0^{(n-1)}$. We thus obtain the solution satisfying conditions (5). If the right-hand side of equation (2) is a many-valued function, there will be several solutions of equation (7) corresponding to initial conditions (5). Every solution not included in the family of the general *solution,* i.e. not obtainable from (6) for any values of constants C_s, is called a singular solution of the equation.

The remarks made in [7] in connection with first order equations must be borne in mind as regards the concepts of general *solution*

and singular solutions. These concepts have to be related to the existence and uniqueness theorem.

If the right-hand side of equation (2) is expanded into a series arranged in positive integral powers of the differences:

$$(x - x_0),\ (y - y_0),\ (y' - y_0'), \ldots, (y^{(n-1)} - y_0^{(n-1)}),$$

on the hypothesis that the absolute values of these differences do not exceed a certain positive number, the solution satisfying initial conditions (5) can be represented as a series

$$y_0 + \frac{y_0'}{1!}(x - x_0) + \frac{y_0''}{2!}(x - x_0)^2 + \ldots \tag{8}$$

for all x sufficiently near x_0. Here, equation (2) itself gives fully defined values of the coefficients of the series, as in the case of first order equations [5]. In fact, on substituting $x = x_0$ and initial conditions (5) in the equation, $y_0^{(n)}$ can be found. We then differentiate (2) with respect to x, substitute $x = x_0$ and initial conditions (5) and $y^{(n)} = y_0^{(n)}$, and thus find $y_0^{(n+1)}$, and so on.

Another procedure can be adopted for finding the coefficients of the series, that of replacing y on both sides of equation (2) by the power series:

$$y = y_0 + \frac{y_0'}{1!}(x - x_0) + \frac{y_0''}{2!}(x - x_0)^2 + \ldots +$$
$$+ \frac{y_0^{(n-1)}}{(x-1)!}(x - x_0)^{n-1} + a_n(x - x_0)^n + a_{n+1}(x - x_0)^{n+1} + \ldots$$

with undetermined coefficients $a_n, a_{n+1}, \ldots$. We arrange the right-hand side of the equation obtained in powers of $(x - x_0)$, then successively determine the coefficients just mentioned by equating the terms in like powers of $(x - x_0)$ on both sides of our identity [5].

Example. We consider the motion of a particle of mass m along a straight line under the action of an elastic force tending to pull the particle back to its position of equilibrium and proportional to the displacement of the particle from this position. We further assume that the motion takes place in a medium whose resistance is expressed as the sum of two terms: the first directly proportional to the velocity, and the second proportional to the cube of the velocity. If we let x denote the displacement of the particle from its equilibrium position, we get the differential equation:

$$mx'' = -k_1 x - k_2 x' - k_3 x'^3,$$

where k_1, k_2, k_3 are positive coefficients of proportionality.

We take a numerical example:

$$x'' = -x - 0.1x' - 0.1\,x'^3 \tag{9}$$

and we look for the solution satisfying the initial conditions:

$$x|_{t=0} = x_0 = 1; \quad x'|_{t=0} = x_0' = 1, \tag{10}$$

as a series arranged in powers of t. We differentiate equation (9) with respect to t:

$$\left.\begin{aligned}
x''' &= -x' - 0.1x'' - 0.3x'^2 x'' \\
x^{(\mathrm{iv})} &= -x'' - 0.1x''' - 0.3\,(x'^2 x''' + 2x' x''^2) \\
x^{(\mathrm{v})} &= -x''' - 0.1x^{(\mathrm{iv})} - 0.3\,(6x'\,x''\,x''' + x'^2\,x^{(\mathrm{iv})} + 2x''^3) \\
x^{(\mathrm{vi})} &= -x^{(\mathrm{iv})} - 0.1x^{(\mathrm{v})} - 0.3\,(12x''^2\,x''' + 6x'\,x'''^2 + 8x'\,x''\,x^{(\mathrm{iv})} + x'^2\,x^{(\mathrm{v})}).
\end{aligned}\right\} \tag{11}$$

We substitute the initial values (10) in equations (9) and (11), and successively compute the initial values of the derivatives:

$$x_0 = 1; \quad x_0' = 1; \quad x_0'' = -1.2; \quad x_0''' = -0.52; \quad x_0^{(\mathrm{iv})} = 0.544;$$
$$x_0^{(\mathrm{v})} = 0.2160; \quad x_0^{(\mathrm{vi})} = 3.1453.$$

On applying Taylor's formula, we get an approximate expression x_1 for the required solution:†

$$x_1 = 1 + t - 0.6t^2 - 0.0867t^3 + 0.0227t^4 + 0.0018t^5 + 0.0044t^6,$$

$$x_1' = 1 - 1.2t - 0.26t^2 + 0.907t^3 + 0.0000t^4 + 0.0262t^5,$$

$$x_1'' = -1.2 - 0.52t + 0.272t^2 + 0{,}036t^3 + 0.1311t^4,$$

which gives a good degree of accuracy for t near zero.

14. Graphical methods of integrating second order differential equations. There is a corresponding curve for every solution of a differential equation of the nth order, and, as in the case of first order equations, we shall call the curve an integral curve of the equation. In the case of a first order differential equation, there was a corresponding tangent field [5].

We now explain the geometrical significance of the second order equation

$$y'' = f(x, y, y'). \tag{12}$$

---

† It is to be noted that we obtain the series for x_1' and x_1'', not by differentiating the series for x_1, but by applying Taylor's formula to x_1' and x_1'':

$$x_1' = x_0' + \frac{x_0''}{1}\,t + \frac{x_0'''}{2!}\,t^2 + \frac{x_0^{(\mathrm{iv})}}{3!}\,t^3 + \frac{x_0^{(\mathrm{v})}}{4!}\,t^4 + \frac{x_0^{(\mathrm{vi})}}{5!}\,t^5,$$

$$x_1'' = x_0'' + \frac{x_0'''}{1}\,t + \frac{x_0^{(\mathrm{iv})}}{2!}\,t^2 + \frac{x_0^{(\mathrm{v})}}{3!}\,t^3 + \frac{x_0^{(\mathrm{vi})}}{4!}\,t^4.$$

Let s be the length of arc of the integral curve, and let α be the angle that the positive direction of the tangent forms with the positive direction of OX. We have [I, 70]:

$$\frac{dy}{dx} = \tan\alpha; \quad \frac{dx}{ds} = \cos\alpha,$$

and we obtain, on differentiating with respect to x:

$$\frac{d^2y}{dx^2} = \frac{1}{\cos^2\alpha} \cdot \frac{d\alpha}{dx} = \frac{1}{\cos^2\alpha}\,\frac{d\alpha}{ds} \cdot \frac{ds}{dx} = \frac{1}{\cos^3\alpha} \cdot \frac{d\alpha}{ds};$$

but $d\alpha/ds$ is the curvature of the curve, as we know from [I, 71]

$$\frac{d\alpha}{ds} = \frac{1}{R}, \tag{13}$$

and the previous equation gives us:

$$\frac{1}{R} = \cos^3\alpha \cdot \frac{d^2y}{dx^2}. \tag{14}$$

We take R positive here, if α increases with increasing s, and negative if α decreases with increasing s.

We take, say, OX directed to the right, and OY directed upwards (Fig. 15). With this, if $R > 0$, the curve rises from right to left with increasing s (counter-clockwise), and in the opposite direction if $R < 0$.

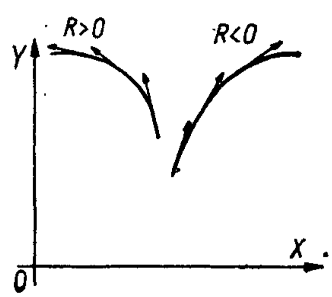

FIG. 15

By (14), the differential equation (12) can be rewritten in the form:

$$\frac{1}{R} = f(x, y, \tan\alpha)\cos^3\alpha. \tag{15}$$

It is clear from this that a *differential equation of the second order gives the radius of curvature, if the position of the point and the direction of the tangent at this point are given.*

This fact gives rise to the method of approximating to the integral curve of a second order equation by means of a curve *with a continuously varying tangent and composed of the arcs of circles.* This method is analogous to that of approximating to the integral curve of a first order equation by means of a step-line [5].

We take the initial conditions for the required integral curve as:

$$y|_{x=0} = y_0; \quad y'|_{x=0} = y_0'.$$

We mark off the point M_0 with coordinates (x_0, y_0) and draw M_0T_0 through the point with slope $y' = \tan\alpha = y_0'$ (Fig. 16).

Equation (15) gives us the corresponding $R = R_0$. We draw M_0C_0 perpendicular to M_0T_0 and equal in length to R_0, then with C_0 as centre construct a small arc M_0M_1 of a circle of radius R_0.

We notice here that the direction of M_0C_0 is determined by the sign of R_0, by what was said above. If, for instance, $R_0 < 0$, movement must be clockwise along the arc of the circle from M_0 to M_1 (Fig. 16). Let (x_1, y_1) be the co-

ordinates of point M_1 and $\tan \alpha_1$ be the slope of the tangent M_1T_1 to the circle passing through M_1. Equation (15) gives the corresponding $R = R_1$. We construct M_1C_1, equal in length to R_1, and perpendicular to M_1T_1, i.e. lying along the straight line M_1C_0, its direction being determined by the sign of R_1; then with C_1 as centre, we draw a small arc M_1M_2 of radius R_1. We proceed from M_2 as from M_1, i.e. find from (15) the corresponding $R = R_2$, draw the line M_2C_2, equal in length to R_2, etc.

A straight rule is used for the above construction, with a hole for a pencil at one end. The quantity R is measured off on a graduated scale that runs

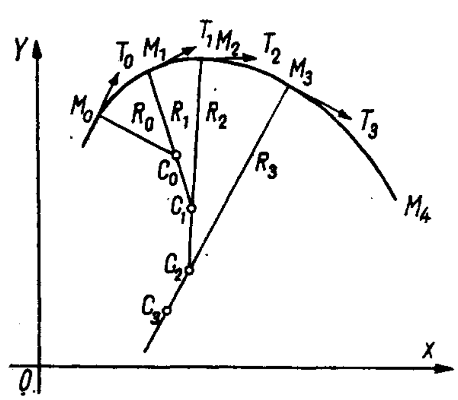

FIG. 16

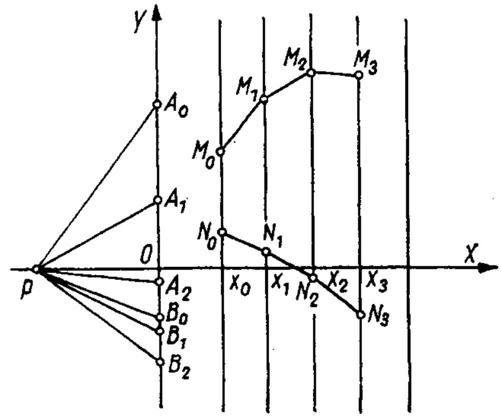

FIG. 17

along the rule from this hole. One leg of a small tripod device is located at the point corresponding to R, whilst the other two legs are on the paper. On shifting the tripod along the scale at points M_1, M_2 etc. in accordance with the variation of R, we do not alter the direction of the tangent at these points; hence we obtain the required curve.

We now give another method of graphical integration of equation (12), providing an approximation to the integral curve in the form of a step line. The method is a generalization of that illustrated in Fig. 9. In addition to y, we introduce the unknown function $z = y'$. We now obtain, in place of the single second order equation (12), a system of two first order equations with two unknown functions y and z:

$$\frac{dy}{dx} = z; \quad \frac{dz}{dx} = f(x, y, z). \tag{16}$$

We apply the method to be explained in the general case of any two first order equations:

$$\frac{dy}{dx} = g(x, y, z); \quad \frac{dz}{dx} = f(x, y, z). \tag{17}$$

We take x as abscissa, and y and z as ordinates in the same coordinate system, so that there will be two integral curves corresponding to every solution of system (17).

We mark off the unit length $\overline{OP}$ along the negative direction of the abscissa (Fig. 17). The values of $f(x, y, z)$ and $g(x, y, z)$ are marked off on the axis OY, using the scale in which $\overline{OP}$ is unity; a different scale to that used for the functions may be used for x, y and z.

Let the solution of system (17) be required which satisfies the initial conditions:

$$y|_{x=x_0} = y_0;$$

$$z|_{x=x_0} = z_0.$$

We draw a series of straight lines, parallel to the y axis:

$$x = x_0;$$

$$x = x_1;$$

$$x = x_2; \ldots$$

We mark off points M_0 and N_0 with coordinates (x_0, y_0) and (x_0, z_0). We take $\overline{OA}_0$ and $\overline{OB}_0$ along the y axis, equal to $g(x_0, y_0, z_0)$ and $f(x_0, y_0, z_0)$ respectively. The lines $\overline{PA}_0$ and $\overline{PB}_0$ will have slopes $g(x_0, y_0, z_0)$ and $f(x_0, y_0, z_0)$, and will therefore give the directions of the tangents to the required integral curves at the initial points M_0 and N_0.

We now draw from these latter points $\overline{M_0M_1}$ and $\overline{N_0N_1}$, parallel to $\overline{PA}_0$ and $\overline{PB}_0$, to their intersections with the line $x = x_1$. Let (x_1, y_1) and (x_1, z_1) be the coordinates of the points of intersection M_1 and N_1. We now mark off $\overline{OA}_1$ and $\overline{OB}_1$ on the ordinate axis, equal in length to $g(x_1, y_1, z_1)$ and $f(x_1, y_1, z_1)$.

From points M_1 and N_1 we draw $\overline{M_1M_2}$ and $\overline{N_1N_2}$, parallel to $\overline{PA}_1$ and $\overline{PB}_1$, to their intersections with $x = x_2$ and so on. We thus obtain two step lines $M_0M_1M_2 \ldots$ and $N_0N_1N_2 \ldots$, representing approximations to the required integral curves.

The construction is simplified in the case of system (16), since $g(x, y, z)$ coincides with the ordinate z of the second line $N_0N_1N_2 \ldots$ The second line here gives an approximate graphical representation of the first derivative y'.

The construction is greatly simplified if the differential equation has the form:

$$y'' = f_1(x) + f_2(y) + f_3(y'),$$

which is often encountered in the investigation of the vibrations of material systems with one degree of freedom.

The equation written is equivalent to the system:

$$\frac{dy}{dx} = z;$$

$$\frac{dz}{dx} = f_1(x) + f_2(x) + f_3(z).$$

If the graphs of the functions f_1, f_2 and f_3 are drawn with the same ordinate scale, we can determine $f(x, y, z)$ by simple addition of the ordinates of these three curves for selected corresponding values of the abscissae x, y, and z.

The method described can also be used for systems of n equations of the first order with n unknown functions. We remark that it is sometimes more con-

venient to mark off the unit vector which we denoted by OP, as also the values of the functions $g(x, y, z)$ and $f(x, y, z)$, from some other point O_1 of axis OY, instead of from the origin O. This is done so as to avoid the lines PA_0, PB_0, $\ldots$, giving the directions of the step line, intersecting with the step line itself.

Figure 18 illustrates the construction of the solution of equation (9), satisfying the initial conditions (10).

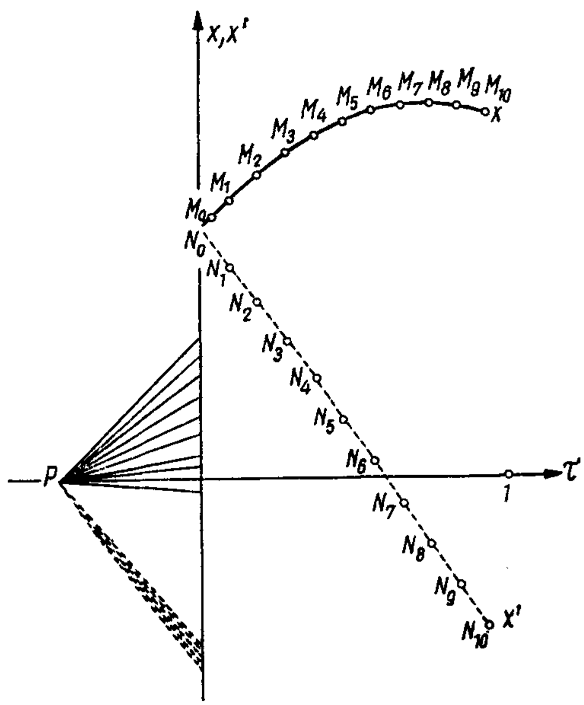

FIG. 18

15. The equation $y^{(n)} = f(x)$ The equation:

$$y^{(n)} = f(x) \tag{18}$$

is a direct generalization of the equation $y' = f(x)$. We start by deriving the formula for the general solution of equation (18). Let $y_1(x)$ be any solution of (18), i.e.:

$$y_1^{(n)}(x) = f(x). \tag{19}$$

We introduce a new required function z in place of y, given by:

$$y = y_1(x) + z. \tag{20}$$

Substitution in (18) gives us the equation for z:

$$y_1^{(n)} + z^{(n)} = f(x),$$

or, using identity (19):

$$z^{(n)} = 0.$$

Since the nth derivative of z must vanish, function z itself is a polynomial of degree $(n - 1)$ with arbitrary constant coefficients:

$$z = C_1 + C_2 x + \ldots + C_n x^{n-1},$$

and (20) gives the general integral of equation (18):

$$y = y_1(x) + C_1 + C_2 x + \ldots + C_n x^{n-1},$$

i.e. *the general solution of equation* (18) *is the sum of any particular solution of the equation and a polynomial of degree* $(n - 1)$ *with arbitrary coefficients.*

It remains for us to find a particular solution of equation (18). We shall seek the solution satisfying the zero initial conditions:

$$\left.\begin{aligned} y\,|_{x=x_0} &= 0\,; \\ y'\,|_{x=x_0} &= 0\,; \\ \ldots y^{(n-1)}\,|_{x=x_0} &= 0\,. \end{aligned}\right\} \qquad (21)$$

On integrating both sides of equation (18) from x_0 to the variable limit x, we obtain:

$$y^{(n-1)} - y_0^{(n-1)} = \int_{x_0}^{x} f(x)\,\mathrm{d}x,$$

where $y_0^{(n-1)}$ is the value of $y_0^{(n-1)}$ for $x = x_0$.

We have $y_0^{(n-1)} = 0$ by the last of conditions (21), so that:

$$y^{(n-1)} = \int_{x_0}^{x} f(x)\,\mathrm{d}x.$$

We obtain $y_0^{(n-2)}$ by again integrating the right-hand side of this equation with respect to x between the limits x_0 and x, and by proceeding in this way, we finally obtain the required function after the nth integration. We usually write this iterated integration as:

$$y = \int_{x_0}^{x} \mathrm{d}x \int_{x_0}^{x} \mathrm{d}x \ldots \int_{x_0}^{x} \mathrm{d}x \int_{x_0}^{x} f(x)\,\mathrm{d}x. \qquad (22)$$

The n times repeated quadrature can be replaced by a single quadrature, as we shall now show.

We expand $y(x)$ by Taylor's formula, with the integral form of remainder term [I, 126]:

$$y(x) = y_0 + (x - x_0)\frac{y_0'}{1!} + (x - x_0)^2\frac{y_0''}{2!} + \ldots +$$

$$+ (x - x_0)^{n-1}\frac{y_0^{(n-1)}}{(n-1)!} + \frac{1}{(n-1)!}\int_{x_0}^{x} (x - t)^{n-1} y^{(n)}(t)\,\mathrm{d}t,$$

where $y_0, y'_0, y''_0, \ldots, y_0^{(n-1)}$ are the values of y and its derivatives for $x = x_0$, whilst t simply denotes the variable of integration. By initial conditions (21):

$$y_0 = y'_0 = y''_0 = \ldots = y_0^{(n-1)} = 0\,,$$

whilst $y^{(n)}(t) = f(t)$ by differential equation (18); hence Taylor's formula above gives:

$$y(x) = \frac{1}{(n-1)!}\int_{x_0}^{x}(x-t)^{n-1} f(t)\,\mathrm{d}t\,. \tag{23}$$

Formula (23) *gives the solution of equation* (18) *for the zero initial conditions* (21), *or, what comes to the same thing, gives an expression for the repeated integral* (22) *in the form of a single integral.*

We get the general solution of equation (18) by adding a polynomial of degree $(n-1)$ with arbitrary coefficients to solution (23). We notice that x appears as the upper limit of integration, as well as under the integral sign, on the right-hand side of (23). Integration is carried out with respect to t, x being meantime considered constant. Formula (23) is obviously also valid for $n = 1$, provided we take $0! = 1$.

16. Bending of a beam. We consider an elastic, prismatic beam, bending under the influence of external forces that may be both concentrated anp continuously distributed (weight, loading).

We take OX along the neutral axis of the beam in its undeformed state, and OY vertically downwards (Fig. 19). We use the convention that forces acting on the beam are positive if directed downwards. We isolate section N of the beam with abscissa x.

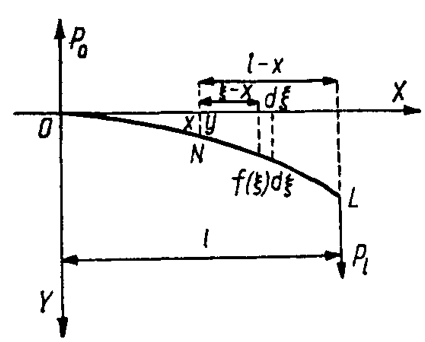

FIG. 19

Let y denote the displacement of the point on the neutral axis, and R the radius of curvature of the deformed axis. It is shown in the theory of strength of materials that, with certain assumption regarding the character of the deformation and the position of the beam relative to axes OX, OY, the equation of equilibrium is to be obtained as follows: we neglect the part of the beam either to the left or to the right of N, and calculate the *bending moment* $M(x)$, equal to the sum of the moments about the neutral line of section N of all the external forces acting on the neglected part, these moments being reckoned positive if, in the case of neglecting the left-hand part, they have a counterclockwise rotation, or in the case of neglecting the right-hand part, they have

a clockwise rotation. The differential equation of the bent axis of the beam now reads:

$$\frac{EI}{R} = M(x)\,, \tag{24}$$

where E is the modulus of elasticity, and I the moment of inertia of the cross-section considered about the neutral line contained in it.

Taking the deformations as in general small, and the axis of the beam on deformation as differing only slightly from axis OX, we can neglect the square of the small quantity y' in the expression for R [I, 71]:

$$R = \frac{(1 + y'^2)^{3/2}}{y''} \sim \frac{1}{y''}\,,$$

which gives, on substituting from equation (24):

$$y'' = \frac{M(x)}{EI}. \tag{25}$$

We now suppose that the only concentrated forces are at the ends of the beam, being equal respectively to P_0 and P_l (in the case of Fig. 19, P_0 is negative); in addition to these, there are bending couples at the ends, the moments of which will be denoted by M_0 and M_l. The distributed loading per unit length of the beam is denoted by $f(x)$.

We find the sum of the moments of the external forces acting on the part NL of the beam (Fig. 19). The loading from any element $d\xi$ with abscissa ξ is $f(\xi)\,d\xi$, and its moment about N is

$$(\xi - x)\, f(\xi)\, d\xi\,,$$

so that the total moment from the full loading of this part is:

$$\int_x^l (\xi - x)\, f(\xi)\, d\xi.$$

On adding the moment of the force P_l, equal to $(l - x)\,P_l$, and the couple of moment M_l, we get:

$$M(x) = \int_0^l (\xi - x)\, f(\xi)\, d\xi + (l - x)\, P_l + M_l. \tag{26}$$

If we calculated with the above sign convention the sum of the moments of all the external forces acting on the part ON of the beam, we should get:

$$M(x) = \int_x^x (x - \xi)\, f(\xi)\, d\xi + x P_0 + M_0. \tag{27}$$

It is easily verified directly that both these expressions are equal. In fact, the equation

$$\int_x^l (\xi - x)\, f(\xi)\, d\xi + (l - x)\, P_l + M_l = \int_0^x (x - \xi)\, f(\xi)\, d\xi + x P_0 + M_0$$

reduces to the following:

$$x\left[\int_0^l f(\xi)\,\mathrm{d}\xi + P_0 + P_l\right] - \left[\int_0^l \xi f(\xi)\,\mathrm{d}\xi + lP_l - M_0 + M_l\right] = 0\,,$$

which is in turn an immediate consequence of the equations:

$$\int_0^l f(\xi)\,\mathrm{d}\xi + P_0 + P_l = 0\,, \tag{28}$$

$$\int_0^l \xi f(\xi)\,\mathrm{d}\xi + lP_l + M_l - M_0 = 0. \tag{29}$$

The first of these expresses the vanishing of the sum of all the external forces, whilst the second equates to zero the sum of the moments about the point O of all the external forces acting on the beam, i.e. they simply express the conditions of equilibrium.

On recalling the expression in [15] for an iterated integral in the form of a simple integral, we can write, by (27):

$$M(x) = \int_0^x \mathrm{d}x \int_0^x f(x)\,\mathrm{d}x + xP_0 + M_0\,, \tag{30}$$

whence

$$\frac{\mathrm{d}M(x)}{\mathrm{d}x} = S(x) = \int_0^x f(\xi)\,\mathrm{d}\xi + P_0, \tag{31}$$

$$\frac{\mathrm{d}^2M(x)}{\mathrm{d}x^2} = f(x). \tag{32}$$

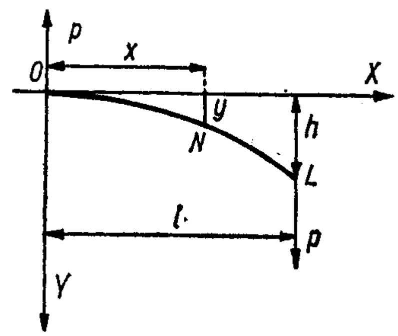

Fig. 20

The quantity $S(x)$, equal to the sum of all the external forces acting to the left of section N, is called *the shear at section* N. Equation (31) shows that *the shear is equal to the derivative of the bending moment.*

Equation (32) has the same form as (25), if we replace in the latter the unknown function y by $M(x)$ and the right-hand side $M(x)/EI$ by $f(x)$. This substitution is of great importance in graphical statics.

Examples. 1. A beam is constrained at the end O and subjected to a concentrated vertical force P at the end L (Fig. 20); the weight of the beam can be neglected. We have in this case:

$$f(x) = 0;\ P_l = P;\ M_l = 0;\ M(x) = (l - x)\,P_l\,,$$

and the equation of equilibrium (25) becomes:

$$y'' = \frac{P}{EI}(l - x).$$

The sag must be zero at the constrained end $x = 0$, and the tangent to the bending axis must coincide with OX, i.e. we have the initial conditions:

$$y\,|_{x=0} = 0 \quad \text{and} \quad y'\,|_{x=0} = 0\,,$$

so that we find [15]:

$$y = \int_0^x (x - \xi)\,\frac{P}{EI}\,(l - \xi)\,\mathrm{d}\xi = \frac{P}{2EI}\left(lx^2 - \frac{x^3}{3}\right).$$

The sag at the end L of the beam is given by:

$$h = y\,|_{x=l} = \frac{Pl^3}{3EI}.$$

The supporting reaction will operate only at end O. Noting that continuous loading is absent here and that $M_l = O$, we have from equations (28) and (29): $R_0 = P_0 = -P$ (reaction force); $M_0 = lP_l$ (reaction couple).

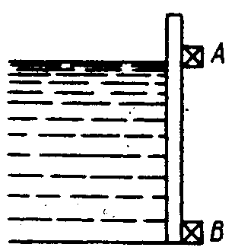

FIG. 21

2. We find the curve of bending of a girder, resting on two supports A and B (Fig. 21) and subjected to a head of water whose level is opposite the upper support (dam). The forces acting on the girder here amount to (1) the continuously distributed head of water, and (2) the reactions at the supports.

Let b be the width of the girder and ϱ be the weight per unit volume of water. If we take a strip of the girder of breadth $\mathrm{d}x$ at a depth x below the level of the water, the head of water at the strip is the weight of a column of water with its base equal to the base-area of the strip and its height equal to the depth of submersion of the strip, i.e.

$$\varrho \cdot b \cdot \mathrm{d}x \cdot x = kx\,\mathrm{d}x \qquad (k = \varrho b).$$

Thus we have in this case: $f(x) = -kx$.

The problem therefore amounts to investigating the bending of a supported beam under the action of continuously distributed loading $f(x) = -kx$.

We start by calculating P_0 and P_l, the reactions of the supports. The total loading is

$$P = \int_0^l k\xi\,\mathrm{d}\xi = \frac{kl^2}{2}.$$

The reactions at the supports O and L due to the elementary loading $k\xi\mathrm{d}\xi$ are, in accordance with the usual law of levers:

$$\frac{k\xi\,(l - \xi)}{l}\,\mathrm{d}\xi \text{ and } \frac{k\xi^2}{l}\,\mathrm{d}\xi.$$

Hence obviously:

$$P_0 = \int_0^l \frac{k\xi\,(l-\xi)}{l}\,\mathrm{d}\xi = \frac{kl^2}{6} = \frac{1}{3}P,\ P_l = P - P_0 = \frac{2}{3}P.$$

We have further, by (26):

$$M\,(x) = -\int_0^l (\xi - x)\,k\xi\,\mathrm{d}\xi + (l - x)\,P_l =$$

$$= -k\int_x^l (\xi - x)\,\xi\,\mathrm{d}\xi + \frac{2}{3}P\,(l - x) = -\frac{k}{6}\,(x^3 - l^2x).$$

Differential equation (25) of bending now becomes:

$$y'' = \frac{-k}{6EI}\,(x^3 - l^2x)\,, \tag{33}$$

with the obvious conditions:

$$y|_{x=0} = 0; \qquad y|_{x=l} = 0. \tag{34}$$

The general solution is:

$$y = \frac{-k}{6EI}\left(\frac{x^5}{20} - \frac{l^2x^3}{6} + C_1x + C_2\right).$$

Constants C_1 and C_2 are found from conditions (34):

$$C_2 = 0; \qquad C_1 = \frac{7}{60}\,l^4,$$

whence finally:

$$y = \frac{-k}{360EI}\,(3x^5 - 10l^2x^3 + 7l^4x).$$

To find the position and value of maximum deflection we put $x = lt$, and re-write the above expression for y as:

$$y = \frac{-kl^5}{360EI}\,(3t^5 - 10t^3 + 7t) \qquad (0 \leqslant t \leqslant 1).$$

The derivative of the polynomial in brackets:

$$15t^4 - 30t^2 + 7$$

has only one zero in the interval (0, 1):

$$t_0 = \sqrt{1 - 2\sqrt{\frac{2}{15}}} \sim 0\cdot519\ldots,$$

which corresponds to a maximum for $|\,y\,|$.

Maximum deflection thus occurs towards the end L and not at the centre, its value being:

$$h = y|_{x=lt_0} = \frac{-kl^5}{360EI}(3t_0^5 - 10t_0^3 + 7t_0) \sim \frac{-kl^5}{360EI} \cdot 2.348 = \frac{-2.348\,Pl^3}{180EI}.$$

17. Lowering the order of a differential equation. We notice a number of particular cases in which the order of an equation can be lowered.

1. Let the function y and a certain number of consecutive derivatives of y: $y', y'', \ldots, y^{(k-1)}$, be excluded from the equation, which has the form:

$$\Phi(x, y^{(k)}, y^{(k+1)}, \ldots, y^{(n)}) = 0.$$

We introduce the new variable $z = y^{(k)}$, thus lowering the order of the equation by k:

$$\Phi(x, z, z', \ldots, z^{(n-k)}) = 0.$$

On finding the general solution of the last equation:

$$z = \varphi(x, C_1, C_2, \ldots, C_{n-k}),$$

we can find y from the equation:

$$y^{(k)} = \varphi(x, C_1, C_2, \ldots, C_{n-k}),$$

which we discussed in [15].

2. If the equation does not contain the independent variable x, i.e. has the form:

$$\Phi(y, y', y'', \ldots, y^{(n)}) = 0,$$

we take y as independent variable and introduce the new function $p = y'$.

If we take p as a function of y, and dependent on x via y, and use the rule for differentiation of a function of a function, we get the following expressions for the derivatives of y with respect to x:

$$y'' = \frac{dp}{dx} = \frac{dp}{dy}p,$$

$$y''' = \frac{d}{dy}\left(\frac{dp}{dx}p\right) = \frac{d}{dy}\left(\frac{dp}{dy}p\right)p = \frac{d^2p}{dy^2}p^2 + \left(\frac{dp}{dy}\right)^2 p,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

and it is clear from these that the order of the equation is $(n-1)$ in the new variables.

If the transformed equation is integrated:

$$p = \varphi(y, C_1, C_2, \ldots, C_{n-1}),$$

the general solution of the given equation can be obtained by a quadrature:

$$\mathrm{d}y = p\,\mathrm{d}x = \varphi(y, C_1, C_2, \ldots, C_{n-1})\,\mathrm{d}x,$$

whence:

$$\int \frac{\mathrm{d}y}{\varphi(y, C_1, C_2, \ldots, C_{n-1})} = x + C_n.$$

One of the arbitrary constants, C_n, appears as an addition to x, which is equivalent to the fact that any integral curve can be displaced parallel to OX.

3. If the left-hand side of the equation:

$$\Phi(x, y, y', \ldots, y^{(n)}) = 0$$

is a homogeneous function [I, 154] of arguments $y, y', \ldots, y^{(n)}$, the introduction of a new function $u(x)$ in place of y, given by the formula

$$y = \mathrm{e}^{\int u\,\mathrm{d}x},$$

results in an equation of order $(n - 1)$ for u. This follows from the obvious formulae:

$$y' = \mathrm{e}^{\int u\,\mathrm{d}x}\, u; \qquad y'' = \mathrm{e}^{\int u\,\mathrm{d}x}\,(u' + u^2); \ldots$$

and from the fact that, after substituting in the left-hand side of the equation, a certain power of the exponential function written above can be taken outside (by the condition of homogeneity) and can then be cancelled out. The arbitrary constant of the integration in the power of e is an arbitrary factor of y.

Examples. **1.** An equation of the form:

$$y'' = f(y) \tag{35}$$

belongs to case 2. It can also be integrated directly. We multiply both side by $2y'\,\mathrm{d}x = 2\mathrm{d}y$:

$$2y'\,y''\,\mathrm{d}x = 2f(y)\,\mathrm{d}y.$$

The left-hand side is obviously the differential of y'^2, and integration gives us:

$$y'^2 = \int_{y_0}^{y} 2f(y)\,\mathrm{d}y + C_1 = f_1(y) + C_1, \text{ whence } \frac{\mathrm{d}y}{\mathrm{d}x} = \sqrt{f_1(y) + C_1}; \tag{36}$$

We separate the variables and integrate:

$$x + C_2 = \int_{y_0}^{y} \frac{\mathrm{d}y}{\sqrt{f_1(y) + C_1}}. \tag{37}$$

If the initial conditions are:

$$y\,|_{x=x_0} = y_0; \quad y'\,|_{x=x_0} = y_0',$$

we obtain, on substituting $x = x_0$, $y = y_0$, $y' = y_0'$ in (36) and (37):

$$C_1 = y_0'^2; \quad C_2 = -x_0,$$

and the required solution becomes:

$$x - x_0 = \int_{y_0}^{y} \frac{\mathrm{d}y}{\sqrt{\int\limits_{y_0}^{y} 2f(y)\,\mathrm{d}y + y_0'^2}}.$$

Let a particle move along the x axis under the action of a force $F(x)$ which depends only on the position of the point. The differential equation of motion is [13]:

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = F(x).$$

Let x_0, v_0 be the initial abscissa and initial velocity of the particle at $t = 0$:

$$x\Big|_{t=0} = x_0; \quad \frac{\mathrm{d}x}{\mathrm{d}t}\Big|_{t=0} = v_0.$$

If we multiply both sides of the equation by $(\mathrm{d}x/\mathrm{d}t)\,\mathrm{d}t$ and integrate, we get:

$$\frac{1}{2}m\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 - \frac{1}{2}mv_0^2 = \int_{x_0}^{x} F(x)\,\mathrm{d}x \quad \text{or} \quad \frac{1}{2}m\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 - \int_{x_0}^{x} F(x)\,\mathrm{d}x = \frac{1}{2}mv_0^2 \tag{38}$$

The first term on the left-hand side, $m(\mathrm{d}x/\mathrm{d}t)^2/2$, consists of the kinetic energy, and the second term $\left[-\int\limits_{x_0}^{x} F(x)\,\mathrm{d}x\right]$ of the potential energy of the moving particle; and it follows from (38) that the sum of the kinetic and potential energies remains constant throughout the motion. We obtain the relationship between x and t by solving (38) with respect to $\mathrm{d}t$ and integrating.

2. If the bending of a beam is too large to allow for the second derivative y'' being taken instead of the curvature [16], we have to take the accurate equation (24) instead of the approximate equation (25). Our problem now amounts to the following: to find the curve whose curvature is a given function of the abscissa,

$$\frac{1}{R} = \varphi(x). \tag{39}$$

This is a second order differential equation:

$$\frac{y''}{(1 + y'^2)^{3/2}} = \varphi(x).$$

On writing $p = y'$, we get a first order differential equation with variables separable:

$$\frac{dp}{(1 + p^2)^{3/2}} = \varphi(x)\, dx\,,$$

and integration gives us:

$$\frac{p}{\sqrt{1 + p^2}} = \int_{x_0}^{x} \varphi(x)\, dx + C_1\,,$$

whence

$$p = \frac{dy}{dx} = \frac{\int_{x_0}^{x} \varphi(x)\, dx + C_1}{\sqrt{1 - [\int_{x_0}^{x} \varphi(x)\, dx + C_1]^2}} = \psi(x), \tag{40}$$

and finally:

$$y = \int_{x_0}^{x} \psi(x)\, dx + C_2.$$

For the case when the beam is supported rigidly at the end $x = 0$ and is subjected to concentrated loading at the other end $x = l$, we have [16]:

$$M(x) = (l - x)\, P; \quad \varphi(x) = \frac{(l - x)P}{EI} = 2k(l - x)\left(k = \frac{P}{2EI}\right).$$

The equation becomes

$$\frac{y''}{(1 + y'^2)^{3/2}} = 2k(l - x),$$

with the initial conditions:

$$y|_{x=0} = 0; \quad y'|_{x=0} = 0.$$

On setting $x_0 = 0$ in (40), we must also set $C_1 = 0$ by the second initial condition, so that we now have:

$$\frac{dy}{dx} = \frac{\int_0^x 2k(l - x)\, dx}{\sqrt{1 - [\int_0^x 2k(l - x)\, dx]^2}} = k\frac{l^2 - (l - x)^2}{\sqrt{1 - k^2[l^2 - (l - x)^2]^2}} =$$

$$= k\frac{x(2l - x)}{\sqrt{1 - k^2 x^2 (2l - x)^2}}\,. \tag{41}$$

We find y by integrating again and using the condition $y|_{x=0} = 0$:

$$y = \int_0^x \frac{x(2l - x)}{\sqrt{1 - k^2 x^2 (2l - x)^2}}\, dx.$$

The integral written is not expressible in terms of elementary functions. The curve corresponding to equation (41) is called an *elastic curve.*

3. We take the equation:

$$x^2 yy'' = (y - xy')^2,$$

both sides of which are homogeneous functions of y, y', y''. On substituting

$$y = e^{\int u\, dx},$$

we get:

$$x^2 (u' + u^2) = (1 - xu)^2,$$

which gives us the linear equation for u:

$$u' + \frac{2}{x} u - \frac{1}{x^2} = 0.$$

Integration of this latter gives:

$$u = x^{-2} (C_1 + x) = C_1 x^{-2} + x^{-1}.$$

On substituting for u in the expression for y, we get:

$$y = e^{-C_1 x^{-1} + \log x + C}$$

or

$$y = C_2 x e^{C_1 x^{-1}},$$

where we have written $C_2 = e^C$ and replaced $(-C_1)$ by C_1.

18. Systems of ordinary differential equations. A system of n first order equations with n unknown functions has the form, on solving with respect to the derivatives:

$$\left.\begin{aligned} \frac{dy_1}{dx} &= f_1(x, y_1, y_2, \ldots, y_n), \\ \frac{dy_2}{dx} &= f_2(x, y_1, y_2, \ldots, y_n), \\ &\cdots\cdots\cdots\cdots\cdots \\ \frac{dy_n}{dx} &= f_n(x, y_1, y_2, \ldots, y_n). \end{aligned}\right\} \tag{42}$$

An existence and uniqueness theorem is valid as in the case of a single equation: *if*

$$f_i(x, y_1, y_2, \ldots, y_n) \quad (i = 1, 2, \ldots, n)$$

are single-valued functions of their arguments, continuous for all x near x_0 and all y_i near $y_i^{(0)}$, and with continuous first order partial derivatives with respect to y_i, there exists one and only one solution of system (42),

$$y_i = \omega_i(x),$$

satisfying the initial conditions:

$$y_1|_{x=x_0} = y_1^{(0)};\quad y_2|_{x=x_0} = y_2^{(0)};\ \ldots;\ y_n|_{x=x_0} = y_n^{(0)}. \tag{43}$$

We can vary the $y_i^{(0)}$ in the initial conditions, so that the general solution of system (42) contains n arbitrary constants. Instead of appearing in the solution as initial values $y_i^{(0)}$, the arbitrary constants can also appear in the general form:

$$y_i = \psi_i(x, C_1, C_2, \ldots, C_n) \quad (i = 1,\ 2, \ldots, n). \tag{44}$$

We obtain particular solutions of system (42) on assigning definite numerical values to the arbitrary constants $C_1, C_2, \ldots, C_n$. To isolate the solution satisfying conditions (43) from this family, we have to determine the arbitrary constants from the equations

$$y_i^{(0)} = \psi_i(x_0, C_1, C_2, \ldots, C_n) \quad (i = 1, 2, \ldots, n) \tag{44$_1$}$$

and substitute the values obtained in (44).

On solving equations (44) with respect to the arbitrary constants, we obtain formulae which give the *general solution of the system* in the form:

$$\varphi_i(x, y_1, y_2, \ldots, y_n) = C_i \quad (i = 1, 2, \ldots, n), \tag{45}$$

with the essential proviso that these equations are soluble with respect to $y_1, y_2, \ldots, y_n$. Any equation of set (45) is called an integral of system (42), and n such integrals have to be found to make up the general solution of the system; thus it follows that equations (45) must be soluble with respect to $y_1, y_2, \ldots, y_n$.

We can re-write system (42) as a series of proportions:

$$dx = \frac{dy_1}{f_1(x, y_1, y_2, \ldots, y_n)} = \frac{dy_2}{f_2(x, y_1, y_2, \ldots, y_n)} = \cdots$$
$$\cdots = \frac{dy_n}{f_n(x, y_1, y_2, \ldots, y_n)}. \tag{46}$$

On multiplying all the denominators by the same factor, we get a function of variables $x, y_1, y_2, \ldots, y_n$ instead of unity in the denominator of the first fraction. If we denote the variables as $x_1, x_2, \ldots, x_n, x_{n+1}$ for the sake of symmetry, the system of differential equations (42) can be written in the form

$$\frac{dx_1}{X_1} = \frac{dx_2}{X_2} = \ldots = \frac{dx_n}{X_n} = \frac{dx_{n+1}}{X_{n+1}}, \tag{47}$$

where $X_1, X_2, \ldots, X_n, X_{n+1}$ are given functions of variables $x_1, x_2, \ldots, x_n, x_{n+1}$. The symmetry of the new form (47) of system (42)

is convenient for later discussions. In particular, which of the $(n+1)$ variables $x_1, x_2, \ldots, x_{n+1}$ is to be taken as independent variable is no longer fixed with (47). The integrals (45) of the system become in the new notation:

$$\varphi_i(x_1, x_2, x_{n+1}) = C_i \; (i = 1, 2, \ldots, n). \tag{48}$$

When the number of the arbitrary constants that appear in the solution (44) is determined, there must be no question of lowering this number. For instance, the three arbitrary constants in the formulae

$$y_1 = (C_1 + C_2)x + C_3; \; y_2 = C_3 x^2; \; y_3 = x^2 + C_3 x + C_1 + C_2$$

can be reduced to two by putting $C_1 + C_2 = C$. The criterion for the impossibility of such reduction and for equations (44) giving the general solution of the system, consists in our being able to satisfy any initial conditions by suitable choice of arbitrary constants, i.e. in that system (44_1) is soluble with respect to $C_1, C_2, \ldots, C_n$ for any choice of the initial values $y_i^{(0)}$ of the required functions. We assume here that the right-hand sides of equations (42) satisfy the conditions mentioned above.

We now turn to a more detailed consideration of the integrals of the system. Suppose that we have k integrals of system (47):

$$\varphi_i(x_1, x_2, \ldots, x_{n+1}) = C_i \; (i = 1, 2, \ldots, k). \tag{49}$$

The functions $\varphi_i(x_1, x_2, \ldots, x_{n+1})$ themselves, and not the equations (49), are sometimes referred to as integrals of the system, i.e. *a function $\varphi(x_1, x_2, \ldots, x_{n+1})$ is called an integral of the system if it becomes a constant on substituting in it any solution of the system.* Of course it is assumed here that $\varphi(x_1, x_2, \ldots, x_{n+1})$ is not itself a constant. Since we can have what initial conditions we please for the solution, the values of this constant can be taken arbitrarily. If we make up an arbitrary function $F(\varphi_1, \varphi_2, \ldots, \varphi_k)$ of the left-hand sides of equations (49), substitution of any solution of the system will make all the φ_i, and therefore the new function, constant, i.e. in addition to integrals (49) we have the integral

$$F(\varphi_1, \varphi_2, \ldots, \varphi_k) = C, \tag{50}$$

where F is an arbitrary function of its arguments. In other words: *an arbitrary function of any integrals of the system is also an integral of the system.* Equation (50) is not a new integral, being a consequence of integrals (49).

Suppose we have n integrals (48) of system (47). They are said to be *independent*, if equations (48) can be solved with respect to any n of the variables $x_1, x_2, \ldots, x_{n+1}$. Such a solution gives us n functions of a single independent variable, i.e. formulae similar to (44), the formulae being solved with respect to the arbitrary constants in the form (48), i.e. n independent integrals (48) of the system are equivalent to the general solution of the system. It can be shown that the condition for the integrals (48) to be independent is equivalent to there being no one integral which is a consequence of the rest in the sense indicated above, or that there exists between the left-hand sides of equations (48) no relationship of the form

$$\Phi(\varphi_1, \varphi_2, \ldots, \varphi_n) = 0,$$

which is an identity with respect to $x_1, x_2, \ldots, x_{n+1}$.

We have given no test in the above by which we might judge whether integrals (48) are independent. Take the case $n = 2$:

$$\varphi_1(x_1, x_2, x_3) = C_1; \; \varphi_2(x_1, x_2, x_3) = C_2. \tag{51}$$

If we recall the implicit function theorem of [I, 159], we can say that a sufficient condition for equations (51) to be soluble with respect to x_2 and x_3 is that the expression

$$\Delta_{x_2, x_3}(\varphi_1, \varphi_2) = \frac{\partial \varphi_1}{\partial x_2}\frac{\partial \varphi_2}{\partial x_3} - \frac{\partial \varphi_1}{\partial x_3}\frac{\partial \varphi_2}{\partial x_2}$$

should differ from zero. Similar statements apply as regards x_3, x_1 and x_1, x_2. Assuming that φ_1 and φ_2 and their first order derivatives are continuous, it can be shown that the necessary and sufficient condition for the independence of integrals (51) is that at least one of the expressions

$$\Delta_{x_2, x_3}(\varphi_1, \varphi_2), \quad \Delta_{x_3, x_1}(\varphi_1, \varphi_2), \quad \Delta_{x_1, x_2}(\varphi_1, \varphi_2)$$

should not be identically zero. We discuss in Volume III the question of the independence of a system of functions with any number of variables.

19. Examples. 1. We take the system:

$$\frac{dx}{xz} = \frac{dy}{yz} = \frac{dz}{-(x^2 + y^2)}. \tag{52}$$

On cancelling out $1/z$ in the equation

$$\frac{dx}{xz} = \frac{dy}{yz}$$

we obtain an equation with separable variables, integration of which gives:

$$\log x = \log y - C, \text{ i.e. } \log \frac{y}{x} = C,$$

which is equivalent to

$$\frac{y}{x} = C_1.$$

We take as the second equation of the system

$$\frac{dx}{xz} = \frac{dz}{-(x^2 + y^2)}$$

and use the solution obtained to substitute in it $y = C_1 x$. We have on cancelling out $1/x$:

$$\frac{dx}{z} = \frac{dz}{-(1 + C_1^2)\, x}, \quad \text{i.e.} \quad (1 + C_1^2)\, x\, dx + z\, dz = 0.$$

Integration gives:

$$(1 + C_1^2)\, x^2 + z^2 = C_2$$

or, on substituting $C_1 = y/x$:

$$x^2 + y^2 + z^2 = C_2.$$

which is the second solution of the system.

The two solutions of the system are therefore:

$$\frac{y}{x} = C_1; \quad x^2 + y^2 + z^2 = C_2. \tag{53}$$

2. The system of differential equations of motion of a material particle of mass m under the action of a given force has the form:

$$m \frac{d^2x}{dt^2} = X; \ m \frac{d^2y}{dt^2} = Y; \ m \frac{d^2z}{dt^2} = Z, \tag{54}$$

where $X, Y, Z,$ the projections of the force on the coordinate axes, are dependent on time, the position of the particle, and its velocity, i.e. on the variables t, x, y, z, x', y', z'.

On taking the derivatives x', y', z' of x, y, z with respect to t as the unknowns, system (54) leads to the system of six first order equations:

$$\frac{dx}{dt} = x'; \ \frac{dy}{dt} = y'; \ \frac{dz}{dt} = z'; \ m \frac{dx'}{dt} = X; \ m \frac{dy'}{dt} = Y; \ m \frac{dz'}{dt} = Z.$$

The general solution of this system contains six arbitrary constants, the determination of which requires the position of the particle and its velocity to be specified at the initial instant.

The following three equations follow from equations (54):

$$m \left(y \frac{d^2z}{dt^2} - z \frac{d^2y}{dt^2} \right) = yZ - zY$$

$$m \left(z \frac{d^2x}{dt^2} - x \frac{d^2z}{dt^2} \right) = zX - xZ$$

$$m \left(x \frac{d^2y}{dt^2} - y \frac{d^2x}{dt^2} \right) = xY - yX,$$

which can obviously be written in the form:

$$\left.\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} m\left(y\frac{\mathrm{d}z}{\mathrm{d}t} - z\frac{\mathrm{d}y}{\mathrm{d}t}\right) &= yZ - zY\\ \frac{\mathrm{d}}{\mathrm{d}t} m\left(z\frac{\mathrm{d}x}{\mathrm{d}t} - x\frac{\mathrm{d}z}{\mathrm{d}t}\right) &= zX - xZ\\ \frac{\mathrm{d}}{\mathrm{d}t} m\left(x\frac{\mathrm{d}y}{\mathrm{d}t} - y\frac{\mathrm{d}x}{\mathrm{d}t}\right) &= xY - yX\,.\end{aligned}\right\} \tag{55}$$

Let the force be centralized, i.e. always pass through some fixed point, called the centre, which we take as origin. Since the projections of a vector are proportional to its direction cosines, and the vector in the present case passes through the origin and the point (x, y, z), we have

$$\frac{X}{x} = \frac{Y}{y} = \frac{Z}{z}\,;$$

the right-hand sides of equations (55) now vanish, and we have the three integrals of system (54):

$$\begin{gathered} m\left(y\frac{\mathrm{d}z}{\mathrm{d}t} - z\frac{\mathrm{d}y}{\mathrm{d}t}\right) = C_1;\quad m\left(z\frac{\mathrm{d}x}{\mathrm{d}t} - x\frac{\mathrm{d}z}{\mathrm{d}t}\right) = C_2;\\ m\left(x\frac{\mathrm{d}y}{\mathrm{d}t} - y\frac{\mathrm{d}x}{\mathrm{d}t}\right) = C_3.\end{gathered} \tag{56}$$

They express the fact, familiar in mechanics, that the areal velocity of the projections of the moving particle on the coordinate planes is constant.

It follows from equations (56) that

$$C_1 x + C_2 y + C_3 z = 0,$$

whence it is evident that the trajectory is a plane curve. The plane of the trajectory is obviously determined by the centre of the force and by the velocity vector at the initial instant.

Now let X, Y, Z be partial derivatives of some function U, depending on x, y, z. We call U the potential of the force, whilst $(-U)$ is the potential energy of the particle:

$$X = \frac{\partial U}{\partial x}\,;\qquad Y = \frac{\partial U}{\partial y}\,;\qquad Z = \frac{\partial U}{\partial z}\,.$$

If we multiply the equations

$$m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = \frac{\partial U}{\partial x}\,;\quad m\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = \frac{\partial U}{\partial y}\,;\quad m\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = \frac{\partial U}{\partial z}$$

by $\mathrm{d}x/\mathrm{d}t$, $\mathrm{d}y/\mathrm{d}t$, $\mathrm{d}z/\mathrm{d}t$ and add, we get:

$$m\left(\frac{\mathrm{d}x}{\mathrm{d}t}\cdot\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \frac{\mathrm{d}y}{\mathrm{d}t}\cdot\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \frac{\mathrm{d}z}{\mathrm{d}t}\cdot\frac{\mathrm{d}^2 z}{\mathrm{d}t^2}\right) = \frac{\mathrm{d}U}{\mathrm{d}t}\,,$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\cdot\frac{m}{2}\left[\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2\right] = \frac{\mathrm{d}U}{\mathrm{d}t}\,,$$

whence we obtain the integral

$$T - U = C, \tag{57}$$

where

$$T = \frac{m}{2}\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right] = \frac{1}{2} mv^2$$

is the *kinetic energy of the particle.*

Equation (57) expresses the fact that the sum of the kinetic energy T and the potential energy $(-U)$ is constant throughout the time of the motion.

3. We consider a system of n particles, inter-related in such a way that the coordinates of any given particle are defined as functions of the independent parameters $q_1, q_2, \ldots, q_k$, and of time t:

$$x_i = \varphi_i(q_1, q_2, \ldots, q_k, t); \quad y_i = \psi_i(q_1, q_2, \ldots, q_k, t); \quad z_i = \omega_i(q_1, q_2, \ldots, q_k, t)$$
$$(i = 1, 2, \ldots, x). \tag{58}$$

Let the system be acted on by forces of potential U, depending only on the position of the particles; then the projections X_i, Y_i, Z_i, of the forces acting on the ith particle on the coordinate axes, are the partial derivatives of U with respect to x_i, y_i, z_i. Let the masses of the particles be $m_1, m_2, \ldots m_n$. By using equations (58), we can write the kinetic energy:

$$T = \sum_{i=1}^{n} \frac{m_i}{2}\left[\left(\frac{dx_i}{dt}\right)^2 + \left(\frac{dy_i}{dt}\right)^2 + \left(\frac{dz_i}{dt}\right)^2\right],$$

and the function U, in terms of parameters $q_1, q_2, \ldots, q_k$, the motion of the system being then defined, as is well known from mechanics, by the following Lagrangian equations:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial q_s'}\right) - \frac{\partial T}{\partial q_s} = \frac{\partial U}{\partial q_s} \qquad (s = 1, 2, \ldots, k) \tag{59}$$

The function T is clearly a second degree polynomial in the derivatives $q_1', q_2', \ldots, q_k'$ of the parameters with respect to time, and (59) consists of k second order equations, which is equivalent to $2k$ first order equations; integration of equations (59) gives us expressions for the q_k as functions of t and of $2k$ arbitrary constants.

Let us suppose that equations (58) do not contain t. Then t will also be absent from T and U. We multiply equations (59) by $q_1', q_2', \ldots, q_k'$ respectively and add:

$$\sum_{s=1}^{k} q_s' \frac{d}{dt}\left(\frac{\partial T}{\partial q_s'}\right) - \sum_{s=1}^{k} q_s' \frac{\partial T}{\partial q_s} = \frac{dU}{dt}. \tag{60}$$

We notice the obvious equality:

$$\sum_{s=1}^{k} q_s' \frac{d}{dt}\left(\frac{\partial T}{\partial q_s'}\right) - \sum_{s=1}^{k} q_s' \frac{\partial T}{\partial q_s} = \frac{d}{dt}\sum_{s=1}^{k} q_s' \frac{\partial T}{\partial q_s'} -$$
$$- \sum_{s=1}^{k} q_s'' \frac{\partial T}{\partial q_s'} - \sum_{s=1}^{k} q_s' \frac{\partial T}{\partial q_s}.$$

In the present case, T is a homogeneous polynomial in q_s' and

$$\sum_{s=1}^{k}{}' q_s' \frac{\partial T}{\partial q_s'} = 2T,$$

by Euler's theorem regarding homogeneous functions [I, 154]. Hence

$$\sum_{s=1}^{k}{}' q_s' \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial q_s'}\right) - \sum_{s=1}^{k}{}' q_s' \frac{\partial T}{\partial q_s} = 2\frac{\mathrm{d}T}{\mathrm{d}t} - \frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\mathrm{d}T}{\mathrm{d}t},$$

and (60) gives us:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\mathrm{d}U}{\mathrm{d}t},$$

whence the integral of (59) is obtained (total energy integral):

$$T - U = C.$$

4. Knowledge of the integral of the differential equations of motion of a system sometimes enables us to solve the problem of the stability of small oscillations of the system about the position of equilibrium. We state the problem mathematically whilst simplifying the discussion by confining ourselves to the case of three unknowns x, y, z, which satisfy the system of differential equations:†

$$\frac{\mathrm{d}x}{\mathrm{d}t} = X; \quad \frac{\mathrm{d}y}{\mathrm{d}t} = Y; \quad \frac{\mathrm{d}z}{\mathrm{d}t} = Z, \tag{61}$$

where X, Y, Z are known functions of x, y, z, and t, vanishing for

$$x = y = z = 0. \tag{62}$$

With this, system (61) has the obvious solution (62), which corresponds to the position of equilibrium. The position of equilibrium [or simply solution (62)] is said to be stable if, for any given positive ε, there exists an η such that, for any solution of system (61) satisfying the initial conditions:

$$x|_{t=0} = x_0; \quad y|_{t=0} = y_0; \quad z|_{t=0} = z_0,$$

we have

$$|x|, \ |y|, \text{ and } |z| < \varepsilon, \tag{63}$$

provided only that

$$|x_0|, \ |y_0| \text{ and } |z_0| < \eta. \tag{64}$$

Let system (61) have an integral

$$\varphi(x, y, z) = C, \tag{65}$$

not containing t, and such that *the function* $\varphi(x, y, z)$ *has a maximum or a minimum for* $x = y = z = 0$. We show that, with this, the position of equi-

† There are six unknowns in the case of the motion of a single material particle.

librium is stable. By changing the sign of φ if necessary, we can assume that it has a minimum; and by adding a constant to φ, we can assume that the minimum is zero.

Function φ now vanishes at the point $x = y = z = 0$ and is positive at all points (x, y, z) near to, but not at, $(0, 0, 0)$. We take a cube δ_ε with centre at the origin and with a side of length 2ε. The continuous function φ is positive at the surface of the cube and therefore attains a least positive value m, so that over all the surface

$$\varphi \geqslant m > 0. \tag{66}$$

We now take a concentric cube δ_η, about the origin, with length of side 2η, such that the inequality is valid within the cube

$$\varphi < m, \tag{67}$$

which is possible since $\varphi(0, 0, 0) = 0$. Let the point (x, y, z) be situated inside cube δ_η at the initial instant, i.e. condition (64) is fulfilled. Inequality (67) will be valid not only at the initial instant, but throughout the time of the motion. By (65), in fact, φ preserves the constant value C during the motion. But given this fact, point (x, y, z) cannot cross the boundary of cube δ_ε during the motion, since inequality (66) must apply at the boundary, which contradicts (67); condition (63) must thus be satisfied for all $t > 0$, which is what we required to prove.

The unknowns x, y, z can be any geometrical or mechanical values, and we only considered them as the coordinates of a point for the sake of clarity of proof. Suppose, for instance, that T and U in equations (59) do not contain time t, so that the total energy integral is valid. Let the equations apply for $q_s = 0$ $(s = 1, 2, \ldots, k)$:

$$\frac{\partial U}{\partial q_1} = \frac{\partial U}{\partial q_2} = \ldots = \frac{\partial U}{\partial q_k} = 0.$$

Equations (59) now have the evident solution:

$$q_s = q_s' = 0, \tag{68}$$

to which the position of equilibrium of the system corresponds. If it also happens that the potential energy $(-U)$ is a minimum for the $q_s = 0$, we can assert that $(T - U)$ is also a minimum for values (68), since T, which cannot be negative, now vanishes, i.e. is also a minimum. Hence we see that the position of equilibrium corresponding to minimum potential energy is stable with respect to the q_s and q_s' (Lagrange–Dirichlet theorem).

20. Systems of equations and equations of higher orders. We consider the relationship between a system of first order differential equations and a single higher order equation. If we have, for example, one differential equation of the third order:

$$y''' = f(x, y, y', y''),$$

we can replace it on writing $y = y_1$, $y' = y_2$, $y'' = y_3$, by a system of three equations of the first order:

$$\frac{dy_1}{dx} = y_2; \ \frac{dy_2}{dx} = y_3; \ \frac{dy_3}{dx} = f(x, y_1, y_2, y_3).$$

We have already carried out a similar substitution in [14]. Likewise, if we are given a system of two second order equations, for instance:

$$y'' = f_1(x, y, y', z, z'); \qquad z'' = f_2(x, y, y', z, z'),$$

where y and z are required as functions of x, we can replace this by a system of four first order equations; here we introduce the four required functions: $y = y_1$; $y' = y_2$; $z = y_3$; $z' = y_4$.

The first system above can be written in the form:

$$\frac{dy_1}{dx} = y_2; \ \frac{dy_2}{dx} = f_1(x, y_1, y_2, y_3, y_4);$$

$$\frac{dy_3}{dx} = y_4; \ \frac{dy_4}{dx} = f_2(x, y_1, y_2, y_3, y_4).$$

Conversely, we show that integration of a system can in general lead to integration of a single higher order equation. We shall only consider the case of a system of three first order equations solved with respect to the derivatives:

$$y_1' = f_1(x, y_1, y_2, y_3); \quad y_2' = f_2(x, y_1, y_2, y_3); \quad y_3' = f_3(x, y_1, y_2, y_3). \tag{69}$$

Let the first equation contain y_2, and let us solve for this:

$$y_2 = \omega_1(x, y_1, y_1', y_3). \tag{70}$$

On substituting in the remaining two equations of the system we shall obtain equations of the form:

$$\frac{\partial \omega_1}{\partial x} + \frac{\partial \omega_1}{\partial y_1} y_1' + \frac{\partial \omega_1}{\partial y_3} y_3' + \frac{\partial \omega_1}{\partial y_1'} y_1'' = \varphi_2(x, y_1, y_1', y_3);$$

$$y_3' = \varphi_3(x, y_1, y_1', y_3).$$

On substituting for y_3' from the second equation in the first, then solving the first equation for y_1'', we get a system of two equations with two required functions y_1 and y_3, of the form:

$$y_1'' = \varphi(x, y_1, y_1', y_3); \quad y_3' = \psi(x, y_1, y_1', y_3). \tag{71}$$

Let the first equation contain y_3; we solve for this and get

$$y_3 = \omega_3(x, y_1, y_1', y_1''); \tag{72}$$

on substituting in the second of equations (71), we obtain a third order equation in y_1, which may be written as:

$$y_1''' F(x, y_1, y_1', y_1''). \tag{73}$$

Suppose that we have managed to integrate this equation:

$$y_1 = \Phi(x, C_1, C_2, C_3).$$

We obtain y_3, on substituting in equation (72); if we then substitute in (70), we obtain y_2, without further integration. If the first of equations (71) does not contain y_3, we already have a second order equation for y_1, and its general solution will contain two arbitrary constants. On substituting this general solution in the second of equations (71), we get a first order equation for y_3, and integration of this introduces a third arbitrary constant. Finally, y_2 is determined from (70) without further integration.

21. Linear partial differential equations. We have so far considered differential equations containing derivatives of functions of a single independent variable. As already mentioned, such equations are called ordinary differential equations. We now consider a class of partial differential equations which is directly related to the theory of systems of ordinary differential equations.

We return to the system of differential equations (47):

$$\frac{dx_1}{X_1} = \frac{dx_2}{X_2} = \ldots = \frac{dx_{n+1}}{X_{n+1}}. \tag{74}$$

An equation

$$\varphi(x_1, x_2, \ldots, x_{n+1}) = C$$

or a function $\varphi(x_1, x_2, \ldots, x_{n+1})$, not identically constant, is called an integral of system (74) if, on substituting in it any solution of the system obtained in accordance with the existence and uniqueness theorem, we obtain a constant.

Thus, let x_1 be the independent variable, and $x_2, x_3, \ldots, x_{n+1}$ be functions of x_1 representing a solution of system (74). A constant must be obtained on substituting these functions in the expression $\varphi(x_1, x_2, \ldots, x_{n+1})$, i.e. the independent variable must go out as a result of substitution; hence the total differential with respect to x_1 must be equal to zero [I, 69]:

$$\frac{\partial\varphi}{\partial x_1} + \frac{\partial\varphi}{\partial x_2}\frac{dx_2}{dx_1} + \frac{\partial\varphi}{\partial x_3}\frac{dx_3}{dx_1} + \ldots + \frac{\partial\varphi}{\partial x_{n+1}}\frac{dx_{n+1}}{dx_1} = 0,$$

or

$$\frac{\partial \varphi}{\partial x_1} dx_1 + \frac{\partial \varphi}{\partial x_2} dx_2 + \ldots + \frac{\partial \varphi}{\partial x_{n+1}} dx_{n+1} = 0. \tag{75}$$

But the differentials dx_s must be proportional to the X_s, since we have substituted a solution of system (74); hence we obtain the following equation for φ, on replacing the dx_s in (75) by the proportional X_s:

$$X_1 \frac{\partial \varphi}{\partial x_1} + X_2 \frac{\partial \varphi}{\partial x_2} + \ldots + X_{n+1} \frac{\partial \varphi}{\partial x_{n+1}} = 0. \tag{76}$$

The function $\varphi(x_1, x_2, \ldots, x_{n+1})$ must satisfy this equation independently of the precise nature of the solution of system (74) that we have substituted in the function. If we take all the solutions of (74), we can give variables $x_1, x_2, \ldots, x_{n+1}$ whatever values we please, in view of the arbitrariness of the initial conditions in the existence and uniqueness theorem, i.e. $\varphi(x_1, x_2, \ldots, x_{n+1})$ must satisfy equation (76) as an identity in $(x_1, x_2, \ldots, x_{n+1})$. Hence we obtain the following theorem.

THEOREM. *If* $\varphi(x_1, x_2, \ldots, x_{n+1}) = C$ *is an integral of system* (74), *the function* $\varphi(x_1, x_2, \ldots, x_{n+1})$ *must satisfy the partial differential equation* (76).

The converse is easily proved.

THEOREM. *If* $\varphi(x_1, x_2, \ldots, x_{n+1})$ *is any solution of equation* (76), $\varphi(x_1, x_2, \ldots, x_{n+1}) = C$ *is an integral of system* (74).

We need only substitute any solution of system (74) in $\varphi(x_1, x_2, \ldots, x_{n+1})$ and take the total differential:

$$d\varphi(x_1, x_2, \ldots, x_{n+1}) = \frac{\partial \varphi}{\partial x_1} dx_1 + \frac{\partial \varphi}{\partial x_2} dx_2 + \ldots + \frac{\partial \varphi}{\partial x_{n+1}} dx_{n+1}.$$

Since a solution has been substituted, the dx_s can be replaced by the proportional X_s by (74), i.e. we write $dx_s = \lambda X_s$, where λ is a coefficient of proportionality. Hence:

$$d\varphi(x_1, x_2, \ldots, x_{n+1}) = \lambda \left(X_1 \frac{\partial \varphi}{\partial x_2} + X_2 \frac{\partial \varphi}{\partial x_2} + \ldots + X_{n+1} \frac{\partial \varphi}{\partial x_{n+1}} \right).$$

But by hypothesis, φ satisfies equation (76) identically in $x_1, x_2, \ldots, x_{n+1}$, so that $d\varphi(x_1, x_2, \ldots, x_{n+1}) = 0$. The expression for the first order differential is independent of whether the variables are independent or not [I, 153]. In the present case, after substituting a solution of the system, φ will be a function of a single independent variable, say x_1; the differential of the function φ was equal to zero,

i.e. the derivative with respect to x_1 (after substitution) is identically zero, so that φ no longer depends on x_1, i.e. is constant. It follows from this that φ is an integral of the system, which is what we required to prove.

The two theorems just proved establish the equivalence of an integral of system (74) and a solution of the partial differential equation (76). If

$$\varphi_1 = C_1;\ \varphi_2 = C_2;\ \ldots;\ \varphi_k = C_k$$

are k integrals of the system, the arbitrary function $F(\varphi_1, \varphi_2, \ldots, \varphi_k)$ is also an integral of the system, as we have seen, and we can therefore assert that *an arbitrary function of any solutions of equation* (76) *is also a solution of the equation.* If

$$\varphi_1(x_1, x_2, \ldots, x_{n+1}) = C_1;\ \ldots;\ \varphi_n(x_1, x_2, \ldots, x_{n+1}) = C_n \quad (77)$$

are n independent integrals of system (74), the arbitrary function $F(\varphi_1, \varphi_2, \ldots, \varphi_n)$ is a solution of equation (76).

This can be verified directly by substituting $\varphi = F(\varphi_1, \varphi_2, \ldots, \varphi_n)$ in equation (76) and noting the fact that functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ satisfy the equation. We do not dwell on the proof of the fact that this is the general solution of equation (76). The following rule is obtained for integrating (76): *to find the general solution of the linear partial differential equation* (76), *we must form the corresponding system of ordinary differential equations* (74) *then obtain n independent integrals* (77) *for the system; the general solution of* (76) *is then*

$$\varphi = F(\varphi_1, \varphi_2, \ldots, \varphi_n),$$

where F is an arbitrary function of its n arguments.

A linear partial differential equation of the form (76) has two characteristics: its coefficients X_i do not contain the required function φ and its free term is zero. The general case of a linear equation has the form

$$Y_1 \frac{\partial \varphi}{\partial x_1} + Y_2 \frac{\partial \varphi}{\partial x_2} + \ldots + Y_n \frac{\partial \varphi}{\partial x_n} + Y_{n+1} = 0, \quad (78)$$

where $Y_1, Y_2, \ldots, Y_{n+1}$ contain $x_1, x_2, \ldots, x_n$ and φ. We seek the family of solutions of equation (78) as the implicit function

$$\omega(x_1, x_2, \ldots, x_n, \varphi) = C, \quad (79_1)$$

where C is an arbitrary constant. By the rule for differentiation of implicit functions:

$$\frac{\partial \varphi}{\partial x_i} = -\frac{\dfrac{\partial \omega}{\partial x_i}}{\dfrac{\partial \omega}{\partial \varphi}};$$

on substituting in (78), we get the equation for ω:

$$Y_1 \frac{\partial \omega}{\partial x_1} + Y_2 \frac{\partial \omega}{\partial x_2} + \ldots + Y_n \frac{\partial \omega}{\partial x_n} + Y_{n+1} \frac{\partial \omega}{\partial \varphi} = 0\,, \qquad (79_2)$$

which has the two characteristics indicated above. We note that the variables $x_1, x_2, \ldots, x_n, \varphi$ can have any values in view of the arbitrariness of C in (79_1), and hence it follows, as above, that equation (79_2) must be satisfied identically with respect to $x_1, x_2, \ldots, x_n, \varphi$. Its solution leads to integration of the corresponding system of ordinary equations. Having found ω, (79_1) gives us φ. It can be shown that, given certain general assumptions regarding the Y_k, all the solutions of equation (78) can be found in this way.

We notice that the general solution of a partial differential equation contains an arbitrary function, whilst only arbitrary constants appear in the general solutions of ordinary differential equations.

We consider linear partial differential equations in more detail in Volume IV, and establish the corresponding existence and uniqueness theorem.

22. Geometrical interpretation. We give a geometrical interpretation of the above theory in the case of three variables. Suppose we have a tangent field in three-dimensional space, i.e. a direction is defined for each point of the space. On taking any system of rectilinear axes, every direction (or tangent) is defined by three numbers, proportional to the direction cosines of the tangent, i.e. the cosines of the angles formed by the tangent with the coordinate axes. Generally speaking, we have different tangents at different points, and the complete tangent field is defined by three functions:

$$u(x, y, z), \quad v(x, y, z), \quad w(x, y, z)\,, \qquad (80)$$

such that the direction cosines of the tangent at a given point (x, y, z) are proportional to magnitudes (80).

We consider the same problem as in the case of a first order equation, that of finding the curves in space whose tangents are those defined by the field. We know from [I, 160] that the direction cosines of a

tangent are proportional to $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$, whilst if two directions coincide, quantities proportional to their direction cosines must themselves be proportional, i.e. we have the following system of differential equations for obtaining the required curves in space:

$$\frac{\mathrm{d}x}{u(x,\ y,\ z)} = \frac{\mathrm{d}y}{v(x,\ y,\ z)} = \frac{\mathrm{d}z}{w(x,\ y,\ z)}. \tag{81}$$

Integration of this system amounts to *finding its two independent integrals*:

$$\varphi_1(x,\ y,\ z) = C_1; \quad \varphi_2(x,\ y,\ z) = C_2, \tag{82}$$

i.e. such that equations (82) are soluble with respect to any two variables. These two equations define a certain curve in space [I, 160]; we obtain a family of integral curves of system (81) on assigning various numerical values to C_1 and C_2. Initial conditions amount to specifying that the required curve should pass through a given point (x_0, y_0, z_0). The arbitrary constants C_1, C_2 are determined by the initial conditions.

We now turn to the geometrical interpretation of the linear partial differential equation. We again take functions (80) as defining a certain tangent field, as above. It is required to find a surface such that, given any point of it, the corresponding tangent plane contains the direction defined by the field at the point. Let the equation of a family of the required surfaces be:

$$\varphi(x,\ y,\ z) = C.$$

From [I, 160], the direction cosines of normals to these surfaces are proportional to $\partial\varphi/\partial x$, $\partial\varphi/\partial y$, $\partial\varphi/\partial z$, whilst the direction of the normal must be perpendicular to the direction defined by magnitudes (80), in order that this latter may lie in the tangent plane. We apply the usual condition for two lines to be perpendicular [I, 160], and obtain a linear partial differential equation for determining φ:

$$u(x, y, z)\frac{\partial\varphi}{\partial x} + v(x, y, z)\frac{\partial\varphi}{\partial y} + w(x, y, z)\frac{\partial\varphi}{\partial z} = 0. \tag{83}$$

The system of ordinary differential equations corresponding to this last equation is (81), so that the general solution of (83) has the form:

$$\varphi = F(\varphi_1, \varphi_2),$$

whilst the general equation of the required surfaces is

$$F(\varphi_1, \varphi_2) = 0, \tag{84}$$

where F is an arbitrary function of its arguments. We do not need to write an arbitrary constant C in view of the arbitrariness of function F, whilst φ_1 and φ_2 are the two independent integrals (82) of system (81). If we make a definite choice of function F, surface (84) will evidently be the locus of the integral curves of system (81) on which the values of the constants in equations (82) are connected by the relationship:

$$F(C_1, C_2) = 0. \tag{85}$$

The solution of equation (83) is generally speaking made precise if we stipulate that the required surface should pass through a given curve in space (L). The stipulation represents initial conditions for partial differential equation (83). The required surface will evidently be composed of the integral curves of system (81) which start from points of the curve (L), i.e. the initial conditions of which are determined by the coordinates of points of (L). We obtain a definite surface in this way, in view of the existence and uniqueness theorem for system (81). This excludes the case when (L) is itself an integral curve of system (81), when the above procedure leads us to (L) itself and not to a surface.

It can be shown that in general an infinite set of surfaces $\varphi = 0$ passes through the curve (L), where φ satisfies equation (83). A detailed discussion will be found in Volume IV.

Let the equation of (L) be given as a set of two equations:

$$\psi_1(x, y, z) = 0; \quad \psi_2(x, y, z) = 0. \tag{86}$$

If we eliminate variables x, y, z from the four equations (82) and (86), we obtain a relationship between C_1 and C_2 which, by (85), also determines the form that function F must take in order that equation (84) may give the required surface passing through curve (86).

23. Examples. 1. We consider the partial differential equation:

$$xz\frac{\partial\varphi}{\partial x} + yz\frac{\partial\varphi}{\partial y} - (x^2 + y^2)\frac{\partial\varphi}{\partial z} = 0. \tag{87}$$

The corresponding system of ordinary differential equations is:

$$\frac{\mathrm{d}x}{xz} = \frac{\mathrm{d}y}{yz} = \frac{\mathrm{d}z}{-(x^2 + y^2)}. \tag{88}$$

We found its two independent integrals in [19] above:

$$\frac{y}{x} = C_1; \quad x^2 + y^2 + z^2 = C_2. \tag{89}$$

The first equation gives a family of planes passing through the z axis, whilst the second gives a family of spheres with centres at the origin. The integral curves of system (88) will be a family of circles lying in these planes with their centres at the origin. The general solution of equation (87) is

$$\varphi = F\left(\frac{y}{x},\quad x^2 + y^2 + z^2\right). \tag{90}$$

where F is an arbitrary function of its two arguments. Let us find the form of F such that the surface

$$F\left(\frac{y}{x},\quad x^2 + y^2 + z^2\right) = 0 \tag{91}$$

passes through the straight line

$$x = 1;\quad y = z. \tag{92}$$

We eliminate x, y and z from equations (89) and (92). The first of equations (89) gives, together with (92):

$$x = 1;\quad y = C_1;\quad z = C_1;$$

substitution in the second of equations (89) now gives the relationship between C_1 and C_2:

$$1 + 2C_1^2 - C_2 = 0, \quad \text{i.e.} \quad F(C_1, C_2) = 1 + 2C_1^2 - C_2.$$

With this form of function F, (91) becomes the equation of the required surface:

$$1 + 2\frac{y^2}{x^2} - (x^2 + y^2 + z^2) = 0 \quad \text{or} \quad x^2 + 2y^2 - x^2(x^2 + y^2 + z^2) = 0.$$

2. Let the tangent field defined by a system of differential equations be such that its direction is the same at all points of space. Let (a, b, c) be numbers proportional to the directioncosines of this fixed direction. The system of differential equations will be:

$$\frac{\mathrm{d}x}{a} = \frac{\mathrm{d}y}{b} = \frac{\mathrm{d}z}{c} \quad \text{or} \quad c\,\mathrm{d}x - a\,\mathrm{d}z = 0;\quad c\,\mathrm{d}y - b\,\mathrm{d}z = 0,$$

which yields at once the two integrals:

$$cx - az = C_1;\quad cy - bz = C_2.$$

The integral curves are obviously parallel straight lines with the fixed direction referred to. The corresponding partial differential equation

$$a\frac{\partial \varphi}{\partial x} + b\frac{\partial \varphi}{\partial y} + c\frac{\partial \varphi}{\partial z} = 0 \tag{93}$$

defines the surfaces $\varphi(x, y, z) = 0$, representing the locus of certain of these straight lines, i.e. (93) is the equation of certain cylindrical surfaces. Its general solution has the form:

$$\varphi = F(cx - az,\ cy - bz),$$

where F is an arbitrary function, and the general equation of the cylindrical surfaces whose generators have the fixed direction is

$$F(cx - az,\ cy - bz) = 0.$$

3. Let the tangent field be such that its direction at any given point $M(x, y, z)$ coincides with the direction of the radius vector from a fixed point $A(a, b, c)$ to the point $M(x, y, z)$. The projections of the vector on the coordinate axes are

$$x - a, \quad y - b, \quad z - c$$

and these quantities are therefore proportional to the direction cosines of the given direction at M. The corresponding system of differential equations is

$$\frac{\mathrm{d}x}{x - a} = \frac{\mathrm{d}y}{y - b} = \frac{\mathrm{d}z}{z - c}.$$

and we have the two obvious integrals:

$$\frac{x - a}{z - c} = C_1; \quad \frac{y - b}{z - c} = C_2.$$

It is geometrically obvious that the family of straight lines passing through $A(a, b, c)$ is a family of integral curves. The corresponding partial differential equation

$$(x - a)\frac{\partial \varphi}{\partial x} + (y - b)\frac{\partial \varphi}{\partial y} + (z - c)\frac{\partial \varphi}{\partial z} = 0$$

defines conical surfaces with vertex at A, the general equation of these surfaces being

$$F\left(\frac{x - a}{z - c},\ \frac{y - b}{z - c}\right) = 0,$$

where F is an arbitrary function of its two arguments.

We remark that generally only one conical surface can be drawn through a given curve in space (L), generated by the radius vectors from the point A to points of (L). If, however, (L) is one of the integral curves of the system, i.e. is a straight line passing through the point A, an infinite set of conical surfaces can be drawn to contain (L).

4. We take another system of differential equations of the form:

$$\frac{\mathrm{d}x}{cy - bz} = \frac{\mathrm{d}y}{az - cx} = \frac{\mathrm{d}z}{bx - ay}. \tag{94}$$

On equating all three ratios to the differential $\mathrm{d}t$ of a new variable t, we can write:

$$\mathrm{d}x = (cy - bz)\,\mathrm{d}t; \quad \mathrm{d}y = (az - cx)\,\mathrm{d}t; \quad \mathrm{d}z = (bx - ay)\,\mathrm{d}t. \tag{95}$$

Hence two equations are easily obtained, integrable directly. The first equation is obtained by multiplying equations (95) respectively by a, b, c and adding, whilst for the second we multiply equations (95) by x, y, z respectively then add. This gives us the two equations:

$$a\,\mathrm{d}x + b\,\mathrm{d}y + c\,\mathrm{d}z = 0, \quad x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z = 0,$$

integration of which yields the two integrals of the system:

$$ax + by + cz = C_1; \quad x^2 + y^2 + z^2 = C_2. \tag{96}$$

The first integral gives a family of parallel planes, the direction cosines of the normals to which are proportional to the numbers (a, b, c). The second integral gives a family of spheres with centres at the origin. The intersections of these planes and spheres represent the family of integral curves of system (94), which evidently consists of circles lying on the planes and with centres on the straight line

$$\frac{x}{a} = \frac{y}{b} = \frac{z}{c}, \tag{97}$$

which in turn passes through the origin and is perpendicular to all the planes.

It can easily be shown that the corresponding partial differential equation

$$(cy - bz)\frac{\partial \varphi}{\partial x} + (az - cx)\frac{\partial \varphi}{\partial y} + (bx - ay)\frac{\partial \varphi}{\partial z} = 0$$

defines the surfaces of revolution which have (97) as axis of revolution, the general equation of these surfaces being

$$F(ax + by + cz, \quad x^2 + y^2 + z^2) = 0,$$

where F is an arbitrary function of its two arguments. We remark that the form of the denominators in system (97) could be obtained from geometrical considerations by suitably specifying the tangent field as was done in previous examples.

5. The problem of orthogonal trajectories in space leads to a linear partial differential equation. Suppose we are given a family of surfaces

$$\omega(x, y, z) = C, \tag{98}$$

dependent on the parameter C, so that, in general, one and only one surface of the family passes through every point in space. We require to find the surface

$$\varphi(x, y, z) = C_1, \tag{99}$$

which intersects all the surfaces (98) at right angles. The condition that the normals to surfaces (99) and (98) should be perpendicular gives us a linear partial differential equation for the required function φ:

$$\frac{\partial \omega}{\partial x}\frac{\partial \varphi}{\partial x} + \frac{\partial \omega}{\partial y}\frac{\partial \varphi}{\partial y} + \frac{\partial \omega}{\partial z}\frac{\partial \varphi}{\partial z} = 0.$$

The corresponding system of ordinary equations:

$$\frac{\mathrm{d}x}{\dfrac{\partial \omega}{\partial x}} = \frac{\mathrm{d}y}{\dfrac{\partial \omega}{\partial y}} = \frac{\mathrm{d}z}{\dfrac{\partial \omega}{\partial z}} \tag{100}$$

defines the curves, whose tangents at every point are normal to surfaces (98) passing through the point concerned. If

$$\varphi_1(x, y, z) = C_1; \quad \varphi_2(x, y, z) = C_2$$

are two independent integrals of system (100), the equation of the required surfaces has the form:

$$F(\varphi_1, \varphi_2) = 0.$$

CHAPTER II

LINEAR DIFFERENTIAL EQUATIONS. SUPPLEMENTARY REMARKS ON THE THEORY OF DIFFERENTIAL EQUATIONS

§ 3. General theory; equations with constant coefficients

24. Linear homogeneous equations of the second order. The simplest part of the theory of differential equations is that dealing with linear equations; these have received the most detailed treatment and are the most commonly encountered in applications. We dealt with the solution of linear equations of the first order in [4]. We consider linear equations of any order in the present chapter, starting with those of the second order.

An equation of the form

$$P(y) = y'' + p(x)\,y' + q(x)\,y = 0, \tag{1}$$

is called a linear homogeneous equation of the second order, where the left-hand side is denoted by $P(y)$ for brevity.

It follows from the linearity of $P(y)$ with respect to the function y and its derivatives that, given arbitrary constants C, C_1 and C_2,

$$P(Cy) = CP(y); \quad P(C_1 y_1 + C_2 y_2) = C_1 P(y_1) + C_2 P(y_2).$$

If $y = y_1$ is a solution of the equation, $P(y_1) = 0$, and obviously $P(Cy_1) = 0$, so that $y = Cy_1$ is also a solution. Similarly, if y_1 and y_2 are solutions,

$$y = C_1 y_1 + C_2 y_2 \tag{2}$$

is also a solution, with arbitrary constants C_1, C_2. Thus, *further solutions of the linear homogeneous equation* (1) *can be obtained by multiplying*

existing solutions by arbitrary constants and adding. It is obvious that linear homogeneous equations of any order will possess the same property. When we refer below to a solution of equation (1), it will be assumed to differ from the trivial solution $y = 0$.

The existence and uniqueness theorem can be stated very simply for equation (1), as we prove in a later paragraph: *if the functions $p(x)$ and $q(x)$ are continuous in the interval $a < x < b$, and if x_0 is any x belonging to the interval, there exists one and only one solution of equation* (1) *satisfying the initial conditions*

$$y|_{x=x_0} = y_0; \quad y'|_{x=x_0} = y'_0,$$

where y_0 and y'_0 are any given numbers. This solution exists throughout the interval $a < x < b$.

We shall in future consider the solutions of equation (1) with x varying in the interval of continuity of $p(x)$ and $q(x)$. In view of the arbitrariness of x_0, y_0, y'_0 in the existence and uniqueness theorem, equation (1) has no singular solutions.

Two solutions y_1 and y_2 of equation (1) are said to be linearly independent if no identity with respect to x exists of the form

$$\alpha_1 y_1 + \alpha_2 y_1 = 0, \tag{3}$$

where α_1 and α_2 are non-zero constant coefficients. In other words, the linear independence of y_1 and y_2 implies that the ratio y_2/y_1 is not a constant, i.e. that the derivative of the ratio

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{y_2}{y_1}\right) = \frac{y_1 y'_2 - y_2 y'_1}{y_1^2} \tag{4}$$

is not identically zero.

We introduce into the discussion the expression

$$\varDelta(y_1 y_2) = y_1 y'_2 - y_2 y'_1, \tag{5}$$

called *the Wronskian* of the solutions y_1 and y_2. A characteristic of the Wronskian is that:

$$\varDelta(y_1, y_2) = \varDelta_0\, \mathrm{e}^{-\int\limits_{x_0}^{x} p(x)\,\mathrm{d}x}, \tag{6}$$

where $\varDelta_0$ is a constant, equal to the value of $\varDelta(y_1, y_2)$ at $x = x_0$.

We prove this by finding the derivative:

$$\frac{\mathrm{d}\varDelta(y_1, y_2)}{\mathrm{d}x} = y'_1 y'_2 + y_1 y''_2 - y'_2 y'_1 - y_2\, y''_1 = y_1 y''_2 - y_2 y''_1.$$

Since y_1 and y_2 are solutions of equation (1), we can write:

$$y_1'' + p(x)\,y_1' + q(x)\,y_1 = 0; \quad y_2'' + p(x)\,y_2' + q(x)\,y_2 = 0.$$

We multiply the first equation by $(-y_2)$ and the second by y_1 and add:

$$y_1 y_2'' - y_2 y_1'' + p(x)\,(y_1 y_2' - y_2 y_1') = 0,$$

so that

$$\frac{\mathrm{d}\Delta\,(y_1, y_2)}{\mathrm{d}x} + p(x)\,\Delta\,(x_1, y_2) = 0. \tag{7}$$

This is a linear homogeneous equation in Δ, and we obtain (6) at once on applying (31_1) of [4].

It follows from this formula that $\Delta(y_1, y_2)$ *is either identically zero, if the constant* Δ_0 *is zero, or is non-zero for all values of* x, since the exponential function does not vanish. We assume here that $p(x)$ is continuous.

By (6), we can write instead of (4):

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{y_2}{y_1}\right) = \frac{\Delta(y_1, y_2)}{y_1^2} = \Delta_0\,\frac{e^{-\int_{x_0}^{x} p(x)\,\mathrm{d}x}}{y_1^2}\,,$$

and hence it follows that *two solutions* y_1 *and* y_2 *of equation* (1) *are linearly independent when, and only when,* $\Delta(y_1, y_2)$ *differs from zero,* i.e. when $\Delta_0 \neq 0$.

We now show that, if y_1 and y_2 are linearly independent solutions of equation (1), (2) gives us, with suitable choice of constants C_1 and C_2, the solution of (1) satisfying any previously assigned initial conditions:

$$y|_{x=x_0} = y_0; \quad y'|_{x=x_0} = y_0'. \tag{8}$$

Let $y_{10}, y_{20}, y_{10}', y_{20}'$ denote the values of y_1 and y_2 and their derivatives for $x = x_0$. To satisfy the initial conditions (8), we have to determine the C_1 and C_2 in (2) from the system of equations

$$C_1 y_{10} + C_2 y_{20} = y_0; \quad C_1 y_{10}' + C_2 y_{20}' = y_0'.$$

It follows from the linear independence of y_1, y_2 that

$$\Delta_0 = y_{10} y_{20}' - y_{20} y_{10}' \neq 0,$$

so that the system written gives us fully defined values for C_1 and C_2, which proves our assertion.

But by the existence and uniqueness theorem [3], every solution of equation (1) is fully defined by its initial conditions, and we can therefore state the following proposition: *if* y_1 *and* y_2 *are two linearly*

independent solutions of equation (1), *all the solutions of the equation are given by* (2).

The problem of integrating (1) thus reduces to finding two linearly independent solutions. Let y_1 be a solution, and y_2 any other solution. We get by integrating relationship (7):

$$\frac{y_2}{y_1} = \Delta_0 \int e^{-\int\limits_{x_0}^{x} p(x)\,dx} \frac{dx}{y_1^2} \quad \text{or} \quad y_2 = \Delta_0 y_1 \int e^{-\int\limits_{x_0}^{x} p(x)\,dx} \frac{dx}{y_1^2}; \tag{9}$$

thus, if a particular solution of equation (1) is known, its second solution can be found by using (9), where Δ_0 is a constant which can be set equal to unity.

It must be remarked that it proves impossible to find this solution explicitly, or even with the aid of a quadrature, in the general case when $p(x)$ and $q(x)$ are functions of x. We shall see, however, that the solutions are obtainable explicitly in some particular cases, including that when $p(x)$ and $q(x)$ are constants, and not functions of x.

We also give later a method of constructing solutions which is often used in applications, viz., the construction as an infinite series.

25. Non-homogeneous linear equations of the second order. An equation of the form:

$$u'' + p(x)\,u' + q(x)\,u = f(x). \tag{10}$$

is called a non-homogeneous linear equation of the second order.

If $p(x)$, $q(x)$ and $f(x)$ are continuous in an interval $a < x < b$, we have, as will be shown later, exactly the same existence and uniqueness theorem as for the homogeneous equation (1). Below, we shall consider the solutions of equation (10) in the interval of continuity of $p(x)$, $q(x)$ and $f(x)$.

Let $u = u_1$ be a solution of the equation, so that:

$$u_1'' + p(x)\,u_1' + q(x)\,u_1 = f(x). \tag{11}$$

On introducing a new function y instead of u:

$$u = y + u_1. \tag{12}$$

and substituting in (10), we get:

$$\left[y'' + p(x)\,y' + q(x)\,y\right] + \left[u_1'' + p(x)\,u_1' + q(x)\,u_1\right] = f(x),$$

or, by (11),

$$y'' + p(x)\,y' + q(x)\,y = 0. \tag{13}$$

This last equation is called *the homogeneous equation corresponding to equation* (10). If y_1 and y_2 are two linearly independent solutions of (13), we have, by (12) and the proposition of the previous article, the formula:

$$u = C_1 y_1 + C_2 y_2 + u_1,$$

where C_1 and C_2 are arbitrary constants, giving all the solutions of equation (10). The property can be stated thus: *the general solution of a non-homogeneous linear equation of the second order is equal to the sum of the general solution of the corresponding homogeneous equation and any solution of the non-homogeneous equation.*

The above proof is obviously also applicable to non-homogeneous linear equations of any order, so that these possess the same property.

The knowledge of two linearly independent solutions of the homogeneous equation (13) enables us, as we shall now see, to find a particular solution of equation (10), and hence its general solution. We use here the method known as Lagrange's method of varying the arbitrary constants [4].

Let y_1 and y_2 be two linearly independent solutions of (13). The general solution is expressed by (2), as we know.

We shall seek a solution of (10) in the same form, except for taking C_1 and C_2 as required functions of x instead of as constants:

$$u = v_1(x)\, y_1 + v_2(x)\, y_2. \tag{14}$$

Since we have two required functions, and not just one, we can subject $v_1(x)$ and $v_2(x)$ to a further condition, apart from (10). We lay down the following condition:

$$v_1'(x)\, y_1 + v_2'(x)\, y_2 = 0. \tag{15}$$

On differentiating (14) and using (15), we obtain:

$$\begin{array}{r|l} q(x)\cdot & u = v_1(x)\, y_1 + v_2(x)\, y_2 \\ p(x)\cdot & u' = v_1(x)\, y_1' + v_2(x)\, y_2' \\ 1\cdot & u'' = v_1(x)\, y_1'' + v_2(x)\, y_2'' + v_1'(x)\, y_1' + v_2'(x)\, y_2'. \end{array}$$

We substitute in the left-hand side of (10), and get:

$$v_1(x)\left[y_1'' + p(x)\, y_1' + q(x)\, y_1\right] + v_2(x)\left[y_2'' + p(x)\, y_2' + q(x)\, y_2\right] +$$
$$+ v_1'(x)\, y_1' + v_2'(x)\, y_2' = f(x).$$

Bearing in mind that y_1 and y_2 are solutions of homogeneous equation (13), and recalling (15), we have the system of equations

$$v_1'(x)\,y_1 + v_2'(x)\,y_2 = 0; \quad v_1'(x)\,y_1' + v_2'(x)\,y_2' = f(x) \tag{16}$$

for determining $v_1'(x)$ and $v_2'(x)$.

By the linear independence of solutions y_1 and y_2,

$$\Delta(y_1, y_2) = y_1 y_2' - y_2 y_1' \neq 0,$$

so that system (16) fully defines $v_1'(x)$ and $v_2'(x)$. We find $v_1(x)$ and $v_2(x)$ by carrying out the integrations, then substitute in (14) and obtain the solution of equation (10).

26. Linear equations of higher orders. Higher order linear equations possess many of the properties of second order equations. We state these without dwelling on their proof.

An equation of the form

$$y^{(n)} + p_1(x)\,y^{(n-1)} + p_2(x)\,y^{(n-2)} + \ldots + p_{n-1}(x)\,y' + \\ + p_n(x)\,y = 0. \tag{17}$$

is called a linear homogeneous equation of the n-th order.

If $y_1, y_2, \ldots, y_k$ are solutions, the sum

$$C_1 y_1 + C_2 y_2 + \ldots + C_k y_k$$

where $C_1, C_2, \ldots, C_k$ are arbitrary constants, is also a solution. The proof of this is exactly as for second order equations [24].

The statement of the existence and uniqueness theorem is also as for second order equations, with the initial conditions taking the form:

$$y\,|_{x=x_0} = y_0; \quad y'\,|_{x=x_0} = y_0'; \quad \ldots; \quad y^{(n-1)}\,|_{x=x_0} = y_0^{(n-1)}.$$

Solutions $y_1, y_2, \ldots, y_k$ are said to be *linearly independent* if there exists no identity in x of the form:

$$\alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_k y_k = 0$$

with constant coefficients $\alpha_1, \alpha_2, \ldots, \alpha_k$, not all of which are zero.

If $y_1, y_2, \ldots, y_n$ are n linearly independent solutions of the equation, all the solutions are given by the formula:

$$y = C_1 y_1 + C_2 y_2 + \ldots + C_n y_n, \tag{18}$$

where the C_i are arbitrary constants. The solution satisfying the initial conditions given above can be obtained by suitable choice of the C_i.

A non-homogeneous linear differential equation of the n-th order has the form:

$$u^{(n)} + p_1(x)\,u^{(n-1)} + p_2(x)\,u^{(n-2)} + \ldots + \\ + p_{n-1}(x)\,u' + p_n(x)\,u = f(x). \tag{19}$$

If u_1 is a solution of this equation, and $y_1, y_2, \ldots, y_n$ are linearly independent solutions of the corresponding homogeneous equation (17), the general solution of (19) is given by

$$u = C_1 y_1 + C_2 y_2 + \ldots + C_n y_n + u_1,$$

where the C_i are arbitrary constants.

With this, if $y_1, y_2, \ldots, y_n$ are known, the solution of (19) can be found from the expression:

$$u = v_1(x)\,y_1 + v_2(x)\,y_1 + \ldots + v_n(x)\,y_n,$$

where the $v'_i(x)$ are determined from the system of first degree equations:

$$\begin{aligned}
v_1'(x)\,y_1 + v_2'(x)\,y_2 + \ldots + v_n'(x)\,y_n &= 0\\
v_1'(x)\,y_1' + v_2'(x)\,y_2' + \ldots + v_n'(x)\,y_n' &= 0\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\\
v_1'(x)\,y_1^{(n-2)} + v_2'(x)\,y_2^{(n-2)} + \ldots + v_n'(x)\,y_n^{(n-2)} &= 0\\
v_1'(x)\,y_1^{(n-1)} + v_2'(x)\,y_2^{(n-1)} + \ldots + v_n'(x)\,y_n^{(n-1)} &= f(x).
\end{aligned}$$

It may be pointed out, to the reader familiar with determinants, that the necessary and sufficient condition for linear independence is precisely analogous to that given above for second order equations. Let $y_1, y_2, \ldots, y_n$ be solutions of (17) as before. The following nth order determinant is called the Wronskian of these solutions:

$$\Delta(y_1, y_2, \ldots, y_n) = \begin{vmatrix} y_1, & y_2, \ldots, & y_n \\ y_1', & y_2', \ldots, & y_n' \\ y_1'', & y_2'', \ldots, & y_n'' \\ \cdots & \cdots & \cdots \\ y_1^{(n-1)}, & y_2^{(n-1)}, \ldots, & y_n^{(n-1)} \end{vmatrix}$$

and an expression can be derived for this, analogous to (6):

$$\Delta(y_1, y_2, \ldots, y_n) = \Delta_0\, e^{-\int_{x_0}^{x} p_1(x)\,dx},$$

where Δ_0 is the value of Δ at $x = x_0$. It follows from this expression, as above, that Δ is either identically zero, or does not vanish for any value of x. A necessary and sufficient condition for the linear independence of solutions $y_1, y_2, \ldots, y_k$ is that the Wronskian should not be identically zero. With this, the arbitrary constants of (18) are fully defined for any initial conditions. As in

the case of second order equations, the existence and uniqueness theorem gives the solution throughout the interval of continuity of the coefficients $p_1(x), p_2(x), \ldots, p_n(x)$ of the equation.

27. Homogeneous equations of the second order with constant coefficients. Before dealing with equations with constant coefficients, we prove a formula of the differential calculus which will be required later. We are familiar with the formula for the derivative of the function e^{rx}, where r is a real number:

$$(e^{rx})' = re^{rx}.$$

We prove that the same formula applies when r is complex and x is the usual real variable, i.e.

$$(e^{(a+bi)x})' = (a + bi)\, e^{(a+bi)x}.$$

It follows from the definition of an exponential function with complex exponent [I, 176] that:

$$e^{(a+bi)x} = e^{ax}(\cos bx + i \sin bx).$$

We get by differentiation in accordance with the usual rules:

$$(e^{(a+bi)x})' = ae^{ax}(\cos bx + i \sin bx) + be^{ax}(-\sin bx + i \cos bx),$$

or, on taking i outside the second bracket and noting that $1/i = -i$,

$$(e^{(a+bi)x})' = ae^{ax}(\cos bx + i \sin bx) + bie^{ax}(\cos bx + i \sin bx) =$$
$$= (a + bi)\, e^{ax}(\cos bx + i \sin bx) = (a + bi)\, e^{(a+bi)x},$$

which is what we wished to prove.

We now turn to the solution of a linear homogeneous equation of the second order with constant coefficients:

$$y'' + py' + qy = 0, \tag{20}$$

where p and q are given numbers. We substitute a function of the form e^{rx} for y in the equation, where r is a real or complex number which we require to find:

$$y = e^{rx}. \tag{21}$$

We get by differentiating and taking e^{rx} outside the bracket:

$$e^{rx}(r^2 + pr + q) = 0,$$

so that (20) will be satisfied if r is a root of the quadratic equation:

$$r^2 + pr + q = 0, \tag{22}$$

this latter being called the *characteristic equation of equation* (20). If the quadratic equation has two distinct roots, $r = r_1$ and $r = r_2$, (21) gives us two linearly independent solutions of the equation:

$$y_1 = e^{r_1 x}, \quad y_2 = e^{r_2 x}. \tag{23}$$

This follows easily from the fact that the ratio $e^{r_1 x} : e^{r_2 x} = e^{(r_1 - r_2)x}$ is not a constant. We now take the case when (22) has equal roots, i.e. when $p^2 - 4q = 0$, the single root of the equation being given here by:

$$r_1 = r_2 = -\frac{p}{2}. \tag{24}$$

Since the method described has led us now only to the one solution $y_1 = e^{r_1 x}$, the other solution remains to be found; this is done by applying the following argument.

We slightly alter the coefficients p and q so that the roots become distinct; for instance, we may let the root r_1 keep its former value (24), whilst the root r_2 is made slightly different. With this, two solutions (23) are obtained. We subtract these two solutions and divide by the constant $(r_2 - r_1)$, which again gives us a solution [24]:

$$y = \frac{e^{r_2 x} - e^{r_1 x}}{r_2 - r_1}. \tag{25}$$

We now let the altered coefficients p and q tend to their original values, for which equation (22) had a double root. With this, r_2 tends to r_1, and both numerator and denominator tend to zero in (25), so that the fraction as a whole tends to a limit equal to the derivative of e^{rx} with respect to r at $r = r_1$. The second solution of the equation is thus $y_2 = xe^{r_1 x}$. Hence, in the case of equal roots of equation (22), we have the following two linearly independent solutions:

$$y_1 = e^{r_1 x}; \quad y_2 = xe^{r_1 x}. \tag{26}$$

We can verify by direct substitution that y_2 is in fact a solution of the equation. The left-hand side of (20) becomes:

$$(r_1^2 xe^{r_1 x} + 2r_1 e^{r_1 x}) + p(r_1 xe^{r_1 x} + e^{r_1 x}) + qxe^{r_1 x} =$$
$$= xe^{r_1 x}(r_1^2 + pr_1 + q) + e^{r_1 x}(2r_1 + p).$$

The first term on the right is zero, since $r = r_1$ is a root of (22), whilst the second term is zero by (24); hence y_2 is a solution of equation (20).

We consider the coefficients p and q as real numbers. But the roots obtained on solving the quadratic equation (22) may be either

real or complex. If (22) has real and distinct roots, (23) gives two linearly independent real solutions, and the general solution of the equation is:

$$y = C_1 e^{r_1 x} + C_2 e^{r_2 x}. \tag{27}$$

Let equation (22) have complex roots; they must be complex conjugates [I, 189], i.e. $r_1 = a + \beta i$ and $r_2 = a - \beta i$, and (23) gives the solutions:

$$y_1 = e^{(a+\beta i)x} = e^{ax}(\cos\beta x + i\sin\beta x);$$
$$y_2 = e^{(a-\beta i)x} = e^{ax}(\cos\beta x - i\sin\beta x).$$

We obtain further solutions by taking the linear combinations of these solutions:

$$\frac{1}{2}(y_1 + y_2) = e^{ax}\cos\beta x, \quad \frac{1}{2i}(y_1 - y_2) = e^{ax}\sin\beta x.$$

These two solutions are also linearly independent, so that in the case when equation (22) has complex roots $r = a + \beta i$, the general solution of the equation is:

$$y = e^{ax}(C_1\cos\beta x + C_2\sin\beta x), \tag{28}$$

Finally, if (22) has equal roots, the general solution is, by (26):

$$y = (C_1 + C_2 x)e^{r_1 x}. \tag{29}$$

We also note the particular case of (28), when equation (22) has pure imaginary roots, i.e. when $a = 0$. We must have $p = 0$ here, whilst q must be a positive number. If we write $q = k^2$, we have roots $\pm ki$ for (22), and the equation

$$y'' + k^2 y = 0 \tag{30}$$

therefore has the general solution:

$$y = C_1\cos kx + C_2\sin kx. \tag{31}$$

28. Non-homogeneous linear equations of the second order with constant coefficients. We now take the non-homogeneous equation

$$y'' + py' + qy = f(x), \tag{32}$$

where p and q are given real numbers as before and $f(x)$ is a given function of x. To find the general solution of this equation, it is sufficient to find any particular solution and add this to the general solution of the corresponding homogeneous equation (20). Since the general

solution of the homogeneous equation is known, the particular solution can be found with the aid of a quadrature by using the method of variation of the arbitrary constants [25]. Let us take as an example an equation of the form:

$$y'' + k^2 y = f(x). \tag{33}$$

The general solution of the corresponding homogeneous equation is given by (31), and we must seek the particular solution of equation (33) in the form:

$$u = v_1(x) \cos kx + v_2(x) \sin kx, \tag{34}$$

where $v_1(x)$ and $v_2(x)$ are required functions of x. Equations (16) now lead to a system of two linear equations for the derivatives of these functions:

$$v_1'(x) \cos kx + v_2'(x) \sin kx = 0$$

$$-v_1'(x) \sin kx + v_2'(x) \cos kx = \frac{1}{k} f(x).$$

Solving these gives us:

$$v_1'(x) = -\frac{1}{k} f(x) \sin kx; \quad v_2'(x) = \frac{1}{k} f(x) \cos kx.$$

We write the primitives as integrals with variable upper limits and with the variable of integration denoted by ξ:

$$v_1(x) = -\frac{1}{k} \int_{x_0}^{x} f(\xi) \sin k\xi \, d\xi; \quad v_2(x) = \frac{1}{k} \int_{x_0}^{x} f(\xi) \cos k\xi \, d\xi,$$

where x_0 is a fixed number. Substitution in (34) gives us the particular solution:

$$u = -\frac{\cos kx}{k} \int_{x_0}^{x} f(\xi) \sin k\xi \, d\xi + \frac{\sin kx}{k} \int_{x_0}^{x} f(\xi) \cos k\xi \, d\xi \tag{34$_1$}$$

or, on taking under the integral sign the factors independent of the variable of integration:

$$u = \frac{1}{k} \int_{x_0}^{x} f(\xi) \sin k(x - \xi) \, d\xi, \tag{34$_2$}$$

and the general solution of (33) becomes:

$$y = C_1 \cos kx + C_2 \sin kx + \frac{1}{k} \int_{x_0}^{x} f(\xi) \sin k(x - \xi) \, d\xi.$$

We notice one point in connection with (34_2). The variable x has a double role on the right-hand side of this expression. Firstly, x is the upper limit of the integral, and secondly, it appears as an additional parameter (and not as variable of integration) under the integral sign, being reckoned constant whilst the integration is carried out. It is easily shown that the particular solution (34_2) satisfies zero initial conditions at $x = x_0$, i.e.

$$u|_{x=x_0} = 0, \quad u'|_{x=x_0} = 0. \tag{34_3}$$

The first of these equalities follows directly from (34_2), since the upper and lower limits of integration coincide at $x = x_0$, and the integral vanishes. The second equality is obtained by finding u' from (34_1), bearing in mind that the derivative of an integral with respect to its upper limit is equal to the integrand at the upper limit. We get after obvious cancelling:

$$u' = \sin kx \int_{x_0}^{x} f(\xi) \sin k\xi \, d\xi + \cos kx \int_{x_0}^{x} f(\xi) \cos k\xi \, d\xi,$$

whence the second equality of (34_3) follows at once.

29. Particular cases. With special forms of the right-hand side of equation (32), the particular solutions can be found much more simply, without recourse to the method of variation of the arbitrary constants. We start by proving a lemma. Let the right-hand side of (32) be the sum of two terms:

$$y'' + py' + qy = f_1(x) + f_2(x), \tag{35}$$

and let $u_1(x)$, $u_2(x)$ be particular solutions of the non-homogeneous equations whose right-hand sides are respectively $f_1(x)$ and $f_2(x)$, i.e.

$$u_1'' + pu_1' + qu_1 = f_1(x); \quad u_2'' + pu_2' + qu_2 = f_2(x).$$

We obtain on adding:

$$(u_1 + u_2)'' + p(u_1 + u_2)' + q(u_1 + u_2) = f_1(x) + f_2(x),$$

so that $(u_1 + u_2)$ is a particular solution of equation (35).

Let us now take a non-homogeneous equation of the form:

$$y'' + py' + qy = ae^{kx}, \tag{36}$$

where a and k on the right-hand side are given numbers. We make use in future of an abbreviated notation for the left-hand side of equation (22), writing

$$\varphi(r) = r^2 + pr + q. \tag{37}$$

We shall seek a solution of (36) in the same form as its right-hand side, i.e. in the form:

$$y = a_1 e^{kx},$$

where a_1 is a required numerical coefficient. On substituting this in (36) and cancelling e^{kx}, we get an equation for a_1 which, by (37), can be written in the form:

$$\varphi(k)\, a_1 = a.$$

If k is not a root of (22), i.e. $\varphi(k) \neq 0$, this last equation gives us a_1. Now let k be a simple root of (22), so that $\varphi(k) = 0$ but $\varphi'(k) \neq 0$ [I. 186]. We shall seek the solution of (33) here in the form:

$$y = a_1\, x e^{kx}.$$

We obtain on substituting in the equation and cancelling e^{kx}:

$$\varphi(k)\, a_1\, x + \varphi'(k)\, a_1 = a,$$

or, since $\varphi(k) = 0$,

$$\varphi'(k)\, a_1 = a,$$

whence we find a_1, since $\varphi'(k) \neq 0$. Finally, if k is a double root of (22), so that $\varphi(k) = \varphi'(k) = 0$, it is easily shown, as above, that the solution of the equation is to be sought in the form:

$$y = a_1\, x^2\, e^{kx}.$$

The same method can be used for finding the solution in the more general case, when the right-hand side has the form $P(x)\, e^{kx}$, where $P(x)$ is a polynomial in x. If k is not a root of equation (22), the solution is to be sought in the form:

$$y = P_1(x)\, e^{kx}, \tag{38}$$

where $P_1(x)$ is a polynomial of the same degree as $P(x)$ and the coefficients of $P_1(x)$ are required to be found. On substituting (38) in the equation, cancelling out e^{kx}, and equating coefficients of like powers of x, we obtain equations for the coefficients of $P_1(x)$.

In the case of k being a root of equation (22), the right-hand side of (38) has to be multiplied by the factor x or x^2, depending on whether k is a simple or a double root of (22).

We now turn to the case of a right-hand side containing trigonometric functions. Let us take the equation, to start with:

$$y'' + py' + qy = e^{kx}\,(a \cos lx + b \sin lx). \tag{39}$$

By using the expressions [I, 177]:

$$\cos lx = \frac{e^{lxi} + e^{-lxi}}{2}, \qquad \sin lx = \frac{e^{lxi} - e^{-lxi}}{2i},$$

we can put the right-hand side of equation (39) in the form:

$$A e^{(k+li)x} + B e^{(k-li)x},$$

where A and B are constants. If the complex numbers $(k \pm li)$ are not roots of equation (22), the solutions must, in accordance with the above, be sought in the form:

$$y = A_1\, e^{(k+li)x} + B_1\, e^{(k-li)x},$$

or, as can be seen on returning from exponential to trigonometric functions via the formula

$$e^{\pm lxi} = \cos lx \pm i \sin lx ,$$

the solution of (39) is to be sought in the form:

$$y = e^{kx}(a_1 \cos lx + b_1 \sin lx) , \tag{40}$$

where a_1 and b_1 are required constants. Similarly, it can be shown that the right-hand side of (40) must be multiplied by x if $(k + li)$ are roots of (22). The constants a_1, b_1 are obtained by substitution of expression (40) in equation (39). We remark that, if say only $\cos lx$ appears in the right-hand side of (39), we still have to take both the $\cos lx$ and the $\sin lx$ terms in the solution (40).

We note a more general result, without dwelling on its proof. If the right-hand side has the form:

$$e^{kx}[P(x) \cos lx + Q(x) \sin lx] ,$$

where $P(x)$ and $Q(x)$ are polynomials in x, the solution must be sought in the same form,

$$e^{kx}[P_1(x) \cos lx + Q_1(x) \sin lx] ,$$

where $P_1(x)$ and $Q_1(x)$ are polynomials in x of degree equal to the greater of the degrees of $P(x)$ and $Q(x)$. If $(k + li)$ are simple roots of equation (22), a factor x must be written in front of the solution.

30. Linear equations of higher orders with constant coefficients. We state without proof in the present section properties of higher order equations analogous to the above. Later on, we explain the general theory of linear equations with constant coefficients by using a special method, known as the *method of symbolic factors*. The present properties will then be proved.

A homogeneous equation of the nth order has the form:

$$y^{(n)} + p_1 y^{n-1} + \ldots + p_{n-1} y' + p_n y = 0 , \tag{41}$$

where $p_1, p_2, \ldots, p_n$ are given real numbers. We write down the characteristic equation, analogous to equation (22):

$$r^n + p_1 r^{n-1} + \ldots + p_{n-1} r + p_n = 0. \tag{42}$$

For every simple real root of this equation, $r = r_1$, there is a corresponding solution $y = e^{r_1 x}$. If the root has multiplicity s, the following s solutions will correspond to it:

$$e^{r_1 x},\ x e^{r_1 x},\ \ldots,\ x^{s-1} e^{r_1 x}.$$

The solutions

$$e^{\alpha x} \cos \beta x \text{ and } e^{\alpha x} \sin \beta x .$$

correspond to a complex conjugate pair of simple roots $r = a \pm \beta i$. If the roots are not simple but have multiplicity s, the following $2s$ solutions correspond to these:

$$e^{ax} \cos \beta x, \; xe^{ax} \cos \beta x, \; \ldots, \; x^{s-1} e^{ax} \cos \beta x$$

$$e^{ax} \sin \beta x, \; xe^{ax} \sin \beta x, \; \ldots, \; x^{s-1} e^{ax} \sin \beta x.$$

In this way, all the roots of equation (42) lead us to solutions of equation (41). On multiplying these solutions by arbitrary constants and adding, we get the general solution of the equation.

To discover a particular solution of the non-homogeneous equation:

$$y^{(n)} + p_1 y^{(n-1)} + \ldots + p_{n-1} y' + p_n y = f(x)$$

we make use of the method of variation of the arbitrary constants [26].

If the right-hand side has the form $P(x) e^{kx}$, where $P(x)$ is a polynomial and k is not a root of equation (42), the solution of the equation can be sought in the form $y = P_1(x) e^{kx}$, where $P_1(x)$ is a polynomial of the same degree as $P(x)$. If k is a root of (42) of multiplicity s, we have to put $y = x^s P_1(x) e^{kx}$. If the right-hand side has the form

$$f(x) = e^{kx} [P(x) \cos lx + Q(x) \sin lx], \tag{43}$$

and $(k \pm li)$ are not roots of equation (42), the solution is to be sought in the same form:

$$y = e^{kx} [P_1(x) \cos lx + Q_1(x) \sin lx],$$

where the degrees of polynomials $P_1(x)$ and $Q_1(x)$ must be taken equal to the greater degree p of polynomials $P(x)$ and $Q(x)$.

On the other hand, if $(k \pm li)$ are roots of (42) of multiplicity s, the factor x^s must be written in front of the right-hand side of the last formula.

Examples. 1. We take the equation

$$y'' - 5y' + 6y = 4 \sin 2x.$$

The corresponding characteristic equation

$$r^2 - 5r + 6 = 0$$

has roots $r_1 = 2$ and $r_2 = 3$. The general solution of the homogeneous equation becomes

$$C_1 e^{2x} + C_2 e^{3x}. \tag{44}$$

The particular solution of the equation is to be sought in the form:

$$y = a_1 \cos 2x + b_1 \sin 2x.$$

We get on substituting in the equation:

$$(2a_1 - 10b_1) \cos 2x + (16a_1 - 4b_1) \sin 2x = 4 \sin 2x,$$

which gives

$$2a_1 - 10b_1 = 0; \quad 16a_1 - 4b_1 = 4,$$

whence $a_1 = 5/19$ and $b_1 = 1/19$, i.e. the particular solution is:

$$y = \frac{5}{19}\cos 2x + \frac{1}{19}\sin 2x.$$

We obtain the general solution of the equation on adding this to (44).

2. We take the fourth order equation:

$$y^{(\mathrm{iv})} - 2y''' + 2y'' - 2y' + y = x \sin x.$$

The corresponding characteristic equation

$$r^4 - 2r^3 + 2r^2 - 2r + 1 = 0$$

can be put in the form:

$$(r^2 + 1)(r - 1)^2 = 0$$

which has a double root $r_1 = r_2 = 1$ and a pair of imaginary conjugates $r_3, r_4 = \pm i$. The general solution of the homogeneous equation becomes:

$$(C_1 + C_2 x)\,\mathrm{e}^x + C_3 \cos x + C_4 \sin x. \tag{45}$$

On comparing the right-hand side with (43), we see that here $k = 0$, $l = 1$, $p = 1$, and $k \pm li = \pm i$ are simple roots of the characteristic equation, so that the particular solution must be sought in the form

$$y = x\left[(ax + b)\cos x + (cx + d)\sin x\right] = (ax^2 + bx)\cos x + (cx^2 + dx)\sin x,$$

where we require to find the coefficients a, b, c, d.

31. Linear equations and oscillatory phenomena. We indicate the importance of linear equations of the second order with constant coefficients in the study of oscillatory phenomena We denote the independent variable by t (time) and the function by x; this notation will often be used in future.

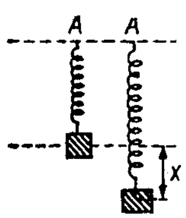

FIG. 22

We consider a body of mass m suspended from a spring and oscillating vertically about its position of equilibrium, where the weight of the body is exactly balanced by the elastic force of the spring.

Let x be the vertical distance of the body from the equilibrium position (Fig. 22). Suppose that the motion takes place in a medium whose resistance is proportional to the velocity $\mathrm{d}x/\mathrm{d}t$.

The following forces will act on the body: (1) the restoring force of the spring, tending to return the body to the equilibrium position, which we shall take as proportional to the displacement x of the body

from the equilibrium position, and (2) the resistive force, proportional to the velocity and acting in the opposite direction. The differential equation of motion becomes:

$$m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = -b\frac{\mathrm{d}x}{\mathrm{d}t} - cx, \quad \text{or} \quad m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + b\frac{\mathrm{d}x}{\mathrm{d}t} + cx = 0.$$

We consider as a second example the motion of a simple pendulum of length l in a medium of resistance proportional to the velocity. The differential equation of motion becomes, as is familiar from mechanics:

$$ml\frac{\mathrm{d}^2 \theta}{\mathrm{d}t^2} = -mg\sin\theta - b\frac{\mathrm{d}\theta}{\mathrm{d}t}, \tag{46}$$

where θ is the angular displacement of the pendulum from the equilibrium position. If the oscillations of the pendulum about the equilibrium position are small, we can take the angle θ for $\sin\theta$, and equation (46) reduces to:

$$ml\frac{\mathrm{d}^2 \theta}{\mathrm{d}t^2} + b\frac{\mathrm{d}\theta}{\mathrm{d}t} + mg\,\theta = 0. \tag{47}$$

If there is an additional external force, depending on time, acting on the pendulum, we get a non-homogeneous equation instead of (47):

$$ml\frac{\mathrm{d}^2 \theta}{\mathrm{d}t^2} + b\frac{\mathrm{d}\theta}{\mathrm{d}t} + mg\,\theta = f(t). \tag{48}$$

The motion in both the above cases is defined by a linear differential equation of the second order with constant coefficients.

We shall write this equation in future in the form:

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + 2h\frac{\mathrm{d}x}{\mathrm{d}t} + k^2 x = 0 \tag{49}$$

or

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + 2h\frac{\mathrm{d}x}{\mathrm{d}t} + k^2 x = f(t). \tag{50}$$

We generally arrive at such an equation when considering the small oscillations of a system with one degree of freedom about its position of equilibrium. The term $2h\,\mathrm{d}x/\mathrm{d}t$ comes from the resistance of the medium or from friction, h being called the coefficient of resistance; the term $k^2 x$ comes from the internal forces of the system, tending to return it to the equilibrium position, k^2 being referred to as the coefficient of restoration; and the term $f(t)$ in equation (50) is due to the external disturbing forces that act on the system. An equation of the type written is encountered, not only in the study

of the oscillations of mechanical systems, but also in other problems of oscillatory phenomena in physics. We take as an example the discharge of a condenser of capacity C through a circuit with resistance R and inductance L. If v is the voltage across the plates of the condenser, we have for the circuit

$$v = Ri + L\frac{di}{dt}, \tag{51}$$

where i is the current in the circuit. In addition, the relationship is known to hold:

$$i = -C\frac{dv}{dt}. \tag{52}$$

Let there also be a source of electromotive force E in the circuit, which we shall take to be positive if it acts in opposition to the direction of i. We have in this case instead of (51):

$$v - E = Ri + L\frac{di}{dt}.$$

On substituting (52) in the equation written, we get the differential equation:

$$LC\frac{d^2v}{dt^2} + RC\frac{dv}{dt} + v = E$$

or

$$\frac{d^2v}{dt^2} + \frac{R}{L}\,\frac{dv}{dt} + \frac{v}{LC} = \frac{E}{LC}. \tag{53}$$

On comparing this equation with equation (50), we see that the term $(R/L)\,dv/dt$ is analogous to the term due to resistance, the term v/LC is analogous to that due to the restoring forces, whilst the term E/LC corresponds with the term from the disturbing force.

If we find v from equation (53), we can also find i, by substituting in (52).

32. Free and forced oscillations. We consider the homogeneous equation

$$x'' + 2hx' + k^2x = 0, \tag{54}$$

corresponding to the case when no external forces are present. The solution of this equation gives the *free*, or *proper*, *vibrations*. The corresponding characteristic equation will be:

$$r^2 + 2hr + k^2 = 0. \tag{55}$$

We split up the further discussion into separate cases.

1. Damped vibrations. In most cases the coefficient of resistance h is fairly small compared with the coefficient of restoration k^2, so that $(h^2 - k^2)$ is negative: $h^2 - k^2 = -p^2$. Equation (55) has in this case conjugate imaginary roots: $r_1, r_2 = -h \pm pi$, and we have for the general solution of (54):

$$x = \mathrm{e}^{-ht}(C_1 \cos pt + C_2 \sin pt). \tag{56}$$

On setting

$$C_1 = A \sin \varphi; \quad C_2 = A \cos \varphi, \tag{57}$$

solution (56) can be written in the form

$$x = A\mathrm{e}^{-ht} \sin (pt + \varphi), \tag{58}$$

or, if we write $p = 2\pi/\tau$,

$$x = A\mathrm{e}^{-ht} \sin \left(\frac{2\pi t}{\tau} + \varphi\right). \tag{59}$$

Here, τ is the period of free vibration, A is the initial amplitude, and φ is the initial phase. If the resistance of the medium is neglected, i.e. we put $h = 0$, the roots of equation (55) are $r = \pm ki$, and we obtain in place of (58):

$$x = A \sin (kt + \varphi). \tag{60}$$

This gives us a *pure harmonic oscillation* of period $\tau = 2\pi/k$. Formula (59) represents damped oscillations [I, 59], the speed of damping being characterized by the factor e^{-ht}. In an interval of time equal to the period, the amplitude decreases in the ratio $\mathrm{e}^{-h\tau}$. The values of the constants C_1 and C_2 in (56), or what amounts to the same thing, constants A and φ in (58), depend on the initial conditions. Suppose the initial conditions are:

$$x|_{t=0} = x_0; \quad x'|_{t=0} = x_0'. \tag{61}$$

On substituting $t = 0$ in (56), we get $C_1 = x_0$. We now differentiate (56) with respect to t:

$$x' = -h\mathrm{e}^{-ht}(C_1 \cos pt + C_2 \sin pt) + p\mathrm{e}^{-ht}(-C_1 \sin pt + C_2 \cos pt),$$

whence we find, on substituting $t = 0$:

$$C_2 = \frac{x_0' + hx_0}{p}, \tag{62}$$

and the solution satisfying initial conditions (61) is finally:

$$x = \mathrm{e}^{-ht}\left(x_0 \cos pt \, \frac{x_0' + hx_0}{p} \sin pt\right) \tag{63}$$

We notice that the coefficient of damping h and the frequency of oscillation $p = \sqrt{k^2 - h^2}$ in solution (63) are completely defined by the coefficients in equation (54). As regards the amplitude A and initial phase φ, which are dependent on the initial conditions, we can write, by (57):

$$A \sin \varphi = x_0; \quad A \cos \varphi = \frac{x_0' + hx_0}{p},$$

from which A and φ may be determined. If $h = 0$, p must be replaced throughout by k.

2. Aperiodic motion. If $(h^2 - k^2)$ is positive:

$$h^2 - k^2 = q^2,$$

the roots of (55) will be:

$$r_1 = -h + q; \quad r_2 = -h - q, \tag{64}$$

and we have [27]:

$$x = C_1 e^{(q-h)t} + C_2 e^{-(q+h)t}. \tag{65}$$

Since we obviously have here $q < h$, both roots of (64) are negative, and x therefore tends to zero on indefinite increase of t.

We differentiate equation (65) with respect to t:

$$x' = C_1 (q - h) e^{(q-h)t} - C_2 (q + h) e^{-(q+h)t}. \tag{66}$$

On putting $t = 0$ in (65) and (66), we get two equations for C_1 and C_2 in terms of the initial conditions (61):

$$C_1 + C_2 = x_0; \quad (q - h) C_1 - (q + h) C_2 = x_0',$$

whence

$$C_1 = \frac{(q + h) x_0 + x_0'}{2q}; \qquad C_2 = \frac{(q - h) x_0 - x_0'}{2q}.$$

3. Special case of aperiodic motion. If, finally, $h^2 - k^2 = 0$, equation (55) has a double root $r_1 = r_2 = -h$, and we get [27]:

$$x = e^{-ht} (C_1 + C_2 t). \tag{67}$$

Since te^{-ht} tends to zero on indefinite increase of t [I, 66], expression (67) also tends to zero.

The non-homogeneous equation

$$x'' + 2hx' + k^2 x = f(t), \tag{68}$$

in which the right-hand side $f(t)$ is due to external forces, defines *forced vibrations.* In the case corresponding to pure harmonic free vibrations, we have:

$$x'' + k^2 x = f(t) \tag{69}$$

and the general solution here is [28]:

$$x = C_1 \cos kt + C_2 \sin kt + \frac{1}{k}\int_0^t f(u) \sin k(t-u)\, \mathrm{d}u,$$

where the last term on the right gives the pure forced vibrations, i.e. the solution of equation (69) satisfying the zero initial conditions:

$$x|_{t=0} = x'|_{t=0} = 0. \tag{70}$$

It can be shown, by using the method of variation of the arbitrary constants, that in the case when the free vibrations are damped, the particular solution of (68) satisfying initial conditions (70) is

$$x_0(t) = \frac{1}{p} \mathrm{e}^{-ht} \int_0^t \mathrm{e}^{hu} f(u) \sin p(t-u)\, \mathrm{d}u, \tag{71}$$

In the aperiodic case, the particular solution becomes:

$$x_0(t) = \frac{1}{2q} \mathrm{e}^{(q-h)t} \int_0^t \mathrm{e}^{(h-q)u} f(u)\, \mathrm{d}u - \frac{1}{2q} \mathrm{e}^{(q+h)t} \int_0^t \mathrm{e}^{(q+h)u} f(u)\, \mathrm{d}u. \tag{72}$$

We leave the proofs to the reader.

33. Sinusoidal external forces and resonance. In practice, the right-hand side is often found to be sinusoidal:

$$x'' + 2hx' + k^2 x = H_0 \sin(\omega t + \varphi_0). \tag{73}$$

We shall seek the solution here in the form of a sinusoidal quantity, of the same frequency ω as the right-hand side [29]:

$$x = N \sin(\omega t + \varphi_0 + \delta_0). \tag{74}$$

We need to define the amplitude N and phase displacement of this vibration. We substitute expression (74) in equation (73):

$$- \omega^2 N \sin(\omega t + \varphi_0 + \delta) + 2h\omega N \cos(\omega t + \varphi_0 + \delta) +$$
$$+ k^2 N \sin(\omega t + \varphi_0 + \delta) = H_0 \sin(\omega t + \varphi_0).$$

We write the argument of the trigonometric functions on the left of this last equation as the sum of the two terms $(\omega t + \varphi_0)$ and δ. We now use the

formulae for the sine and cosine of a sum, and obtain:

$$[(k^2 - \omega^2) N \cos \delta - 2h\omega N \sin \delta] \sin (\omega t + \varphi_0) +$$
$$+ [2h\omega N \cos \delta + (k^2 - \omega^2) N \sin \delta] \cos (\omega t + \varphi_0) = H_0 \sin (\omega t + \varphi_0).$$

On equating the coefficient of $\sin (\omega t + \varphi_0)$ to the constant H_0, and the coefficient of $\cos (\omega t + \varphi_0)$ to zero, we get two equations for N and δ:

$$(k^2 - \omega^2) N \cos \delta - 2h\omega N \sin \delta = H_0; \quad 2h\omega N \cos \delta + (k^2 - \omega^2) N \sin \delta = 0.$$

We solve these with respect to $\cos \delta$ and $\sin \delta$:

$$\cos \delta = \frac{(k^2 - \omega^2) H_0}{N[(k^2 - \omega^2)^2 + 4h^2 \omega^2]}; \quad \sin \delta = -\frac{2h\omega H_0}{N[(k^2 - \omega^2)^2 + 4h^2 \omega^2]}.$$

Squaring both sides of each and adding gives:

$$1 = \frac{H_0^2}{N^2 [(k^2 - \omega^2)^2 + 4h^2 \omega^2]},$$

whence we find

$$N = \frac{H_0}{\sqrt{(k^2 - \omega^2)^2 + 4h^2 \omega^2}}. \tag{75}$$

On substituting this value for N in the above expressions for $\sin \delta$ and $\cos \delta$, we obtain the formulae for δ:

$$\cos \delta = \frac{k^2 - \omega^2}{\sqrt{(k^2 - \omega^2)^2 + 4h^2 \omega^2}}; \quad \sin \delta = -\frac{2h\omega}{\sqrt{(k^2 - \omega^2)^2 + 4h^2 \omega^2}}. \tag{76}$$

Having found N and δ, (74) now gives the sinusoidal part of the solution of equation (73), whilst its general solution becomes

$$x = A e^{-ht} \sin (pt + \varphi) + N \sin (\omega t + \varphi_0 + \delta), \tag{77}$$

where A and φ are arbitrary constants, determined by the initial conditions. We assume here that $h^2 - k^2 = -p^2 < 0$, i.e. that the free vibrations are damped. The first term in (77) rapidly decreases with increasing t, due to the presence of the factor $e^{-ht} (h > 0)$, so that this term only has a noticeable influence on x for t close to zero (transient process); afterwards, x is determined almost exclusively by the second purely sinusoidal term, which is independent of the initial conditions (steady-state process).

We now investigate expressions (75) and (76), which define the amplitude N and the phase difference δ between solution (74) and the right-hand side of equation (73).

If the right-hand side of equation (73) consisted only of the constant H_0, the equation would be

$$x'' + 2hx' + k^2 x = H_0$$

and would have an obvious particular solution, in the form of a constant:

$$\xi_0 = \frac{H_0}{k^2}.$$

This is the *statical deflection*, which would be produced by a constant force. We introduce into the discussion the ratio

$$\lambda = \frac{N}{\xi_0},$$

which gives a measure of the *dynamical susceptibility* of the system to the action of external forces. We get by using (75) and the expression for ξ_0

$$\lambda = \frac{k^2}{\sqrt{(k^2 - \omega^2) + 4h^2\omega^2}} = \frac{1}{\sqrt{\left(1 - \frac{\omega^2}{k^2}\right)^2 + \frac{4h^2}{k^2}\cdot\frac{\omega^2}{k^2}}}.$$

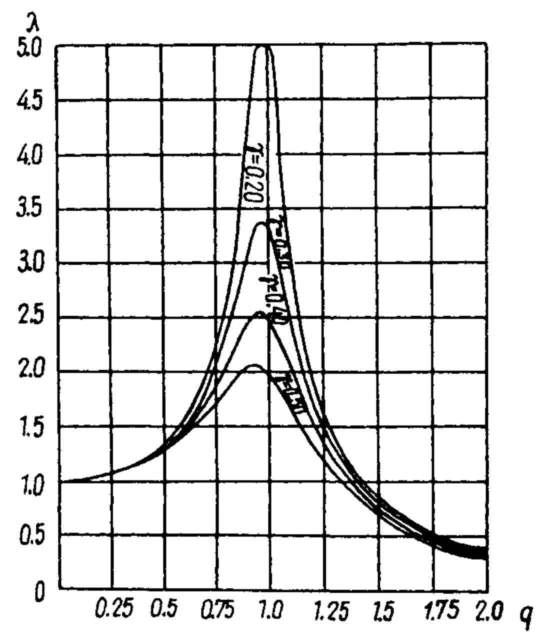

FIG. 23

It is clear from this last expression that λ depends only on the two ratios

$$q = \frac{\omega}{k}, \quad \gamma = \frac{2h}{k}. \tag{78}$$

The mechanical significance of the first ratio may be explained as follows. If there were no resistance, the free vibrations would be given by (60):

$$x = A \sin (kt + \varphi)$$

and the period would be $\tau = 2\pi/k$. Let the period of the disturbing force be denoted by $T = 2\pi/\omega$. We now have for q:

$$q = \frac{\tau}{T}, \tag{79}$$

i.e. q is equal to the ratio of the period of free vibration of the system without resistance to the period of the disturbing force.

We now have for λ:

$$\lambda = \frac{1}{\sqrt{(1-q^2)^2 + \gamma^2 q^2}}, \tag{80}$$

where q is as described above, whilst the constant γ is obviously independent of the action of external forces from its definition. Since h is small, γ is usually small, and if q is not close to unity, the value of λ is approximately $1/(1-q^2)$. Figure 23 illustrates λ as a function of q for various given values of γ.

On dividing numerator and denominator in expressions (76) by k^2, we get the formulae:

$$\left.\begin{aligned} \cos\delta &= (1-q^2)\lambda; \\ \sin\delta &= -\gamma q\lambda, \end{aligned}\right\} \tag{81}$$

These give the phase difference between the external force and the disturbance produced by it.

Since λ depends on q, it is indirectly dependent on the period T of the external force. Let us find the maximum of λ as a function of q. All we need for this is to find the minimum of

$$\frac{1}{\lambda^2} = (1-q^2)^2 + \gamma^2 q^2$$

as a function of q^2. Itis easily shown that the minimum occurs with $q^2 = 1 - \gamma^2/2$, and is equal to $(\gamma^2 - \gamma^4/4)$. Hence it follows that maximum λ occurs with

$$q = \sqrt{1 - \frac{\gamma^2}{2}} \tag{82}$$

and is equal to

$$\lambda_{\max} = \frac{1}{\gamma} \cdot \frac{1}{\sqrt{1 - \frac{\gamma^2}{4}}}.$$

For γ small, the q corresponding to maximum λ is near unity, i.e. the period of the external force which, with a given amplitude, produces the greatest effect is close to the period of the free vibration. The difference between these periods, which depends on γ, is due to the presence of resistance.

If there is no resistance, $\gamma = 0$, and maximum λ is infinity, occurring at $q = 1$.

In this case, characterized by the conditions $h = 0$ and $\omega = k$, equation (73) becomes

$$x'' + k^2 x = H_0 \sin(kt + \varphi_0), \tag{83}$$

and its solution cannot in fact be sought in the form (74).

We suggest that the reader prove that equation (83) has the solution

$$x = -\frac{H_0}{2k} t\cos(kt + \varphi_0),$$

which contains t as a factor [29].

We return to the case when resistance is present, i.e. $h \neq 0$. As is clear from the graph, λ increases rapidly before the maximum, and decreases rapidly

afterwards. This is also easily seen from (80), with small γ. On substituting in (81) $\lambda_{\max}$ and the expression for q of (82), we get:

$$\cos\delta = \frac{\gamma}{2}\,\frac{1}{\sqrt{1-\frac{\gamma^2}{4}}};\quad \sin\delta = -\frac{\sqrt{1-\frac{\gamma^2}{2}}}{\sqrt{1-\frac{\gamma^2}{4}}},$$

whence it is clear that, for greatest effect of the external force, and for small γ, the phase difference δ is close to $(-\pi/2)$.

We now return to (77). Even for fairly small values of t, the first term, giving the free damped vibrations, will be small compared with the second. We shall now vary ω, i.e. the period T of the disturbing force. By what has been said above, the following effect will now be obtained: as T approaches a certain value, the forced vibrations will rapidly increase, attain a maximum, and then rapidly fall off as T passes the value concerned. This phenomenon is called *resonance*. It is encountered in a great variety of processes of an oscillatory nature: in mechanical vibrations, electrical oscillations, sound, etc.

We now suppose that the right-hand side of the equation contains the sum of several sinusoidal quantities:

$$x'' + 2hx' + k^2 x = \sum_{i=1}^{m} H_i \sin(\omega_i t + \varphi_i). \tag{84}$$

For every term on the right, there is a corresponding proper forced vibration of the form

$$N_i \sin(\omega_i t + \varphi_i + \delta_i) \qquad (i = 1, 2, \ldots, m),$$

where N_i and δ_i are given by (75) and (76), if the right-hand side of the equation is known. The sum of the above forced vibrations will correspond to the sum of the external forces, i.e. the particular solution of equation (84) is [29]

$$x = \sum_{i=1}^{m} N_i \sin(\omega_i t + \varphi_i + \delta_i). \tag{85}$$

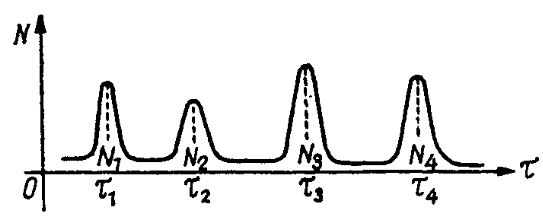

FIG. 24

We now show how the amplitudes and phases of the terms on the right-hand side of equation (84) can be found when they are unknown, by observing the forced vibrations.

Suppose that we are able to vary k^2, i.e. the period τ of the free vibration. The following effect will now occur: as τ approaches a certain value τ_1, the amplitude of the forced vibrations will rapidly increase, reach a maximum, then fall off on further variation of τ; it will now remain small until τ approaches a new value τ_2, corresponding to a second maximum of amplitude of the type just described, and so on.

These maxima are due to resonance with the individual external forces appearing on the right-hand side of equation (84), and the values of $\tau_1, \tau_2, \ldots$ give approximately the periods of the external forces. If we plot the periods of

the free vibrations along the axis of abscissae, and the amplitudes of the forced vibrations on the axis of ordinates, we obtain a curve with several maxima (Fig. 24).

For $\tau = \tau_j$ (or $k = k_j = 2\pi/\tau_j$), the term for which ω_j is near k_j will be large compared with the other terms in the sum (85). By observing experimentally the maximum values of the amplitudes of the forced vibrations, and taking these as approximately equal to the N_j, we can use the formulae:

$$N_j \sim \frac{H_j}{\sqrt{(k_j - \omega_j)^2 + 4h^2 \omega_j^2}},$$

whilst bearing in mind that k_j is close to ω_j, to find the approximate values of the intensities of the forces:

$$H_j \sim 2hk_jN_j.$$

34. Impulsive external forces. We consider forced vibrations without friction,

$$x'' + k^2 x = f(t) \tag{86}$$

and take a special type of external force $f(t)$, acting only in a short interval of time, from $t = 0$ to $t = T$, which rises from zero at the beginning of the interval, reaches a positive maximum, then diminishes to zero (Fig. 25).

The general solution of equation (86) has the form [32]:

$$x = C_1 \cos kt + C_2 \sin kt + \frac{1}{k}\int_0^t f(u) \sin k(t-u)\, du.$$

Let the system be in the equilibrium position with zero initial velocity at $t = 0$:

$$x|_{t=0} = x'|_{t=0} = 0. \tag{87}$$

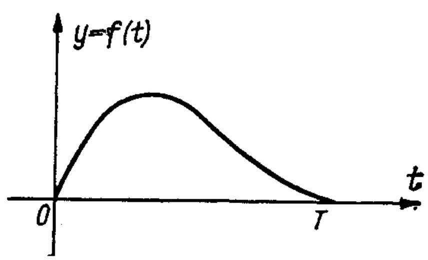

FIG. 25

We know that the particular solution corresponding to these initial conditions is:

$$x = \frac{1}{k}\int_0^t f(u) \sin k(t-u)\, du,$$

which we shall now investigate.

When $t > T$ the integral reduces to the integral over the interval $(0, T)$, since by hypothesis

$$f(u) = 0 \quad \text{for} \quad u > T.$$

Consequently,

$$x = \frac{1}{k}\int_0^T f(u) \sin k(t-u)\, du \quad \text{for} \quad t > T.$$

or

$$x = \frac{1}{k} \sin kt \int_0^T f(u) \cos ku\, du - \frac{1}{k} \cos kt \int_0^T f(u) \sin ku\, du.$$

Since the function $f(u)$ is positive by hypothesis in the interval $(0, T)$, we can apply the mean value theorem [I, 95] to the integral written:

$$\int_0^T f(u) \cos ku \, du = \cos k\theta_1 T \int_0^T f(u)\, du$$

$$(0 < \theta_1 \text{ and } \theta_2 < 1)$$

$$\int_0^T f(u) \sin ku \, du = \sin k\theta_2 T \int_0^T f(u)\, du.$$

Let the duration T of the action of the external force be small compared with the period of free vibration $\tau = 2\pi/k$.

Since $kT = 2\pi T/\tau$ now becomes a small quantity, we can replace $\cos k\theta_1 T$ by unity and $\sin k\theta_2 T$ by zero, so that we get:

$$x = \frac{1}{k} I \sin kt, \tag{88}$$

where

$$I = \int_0^T f(t)\, dt$$

is the magnitude of the impulse of the external force.

It is easily verified that (88) is identical with the formula for the solution of the equation:

$$x'' + k^2 x = 0$$

with the initial conditions [32]:

$$x|_{t=0} = 0; \qquad x'|_{t=0} = I,$$

i.e. *if the action of the external force is of small duration compared with the period of free vibration, the vibration of the system on cessation of the external force will occur as a free vibration with the system driven from its equilibrium position with initial velocity I.*

35. Statical external forces. We now make a different assumption about the force $f(t)$; we let the total interval of action of the force, $(0, T)$, be split into two intervals $(0, T_1)$ and (T_1, T), such that the force is increasing in the first and decreasing in the second sub-interval; and we further suppose that the period of free vibration $\tau = 2\pi/k$ is small compared with the duration of the increase (and decrease) of the force.

We now solve equation (86) with initial conditions (87). We get by integration by parts, and on noting that $f(0) = 0$:

$$\begin{aligned} x &= \frac{1}{k^2} f(u) \cos k(t-u) \Big|_{u=0}^{u=t} - \frac{1}{k^2} \int_0^t f'(u) \cos k(t-u)\, du = \\ &= \frac{1}{k^2} f(t) - \frac{1}{k^2} \int_0^t f'(u) \cos k(t-u)\, du. \end{aligned} \tag{89}$$

The first term $f(t)/k^2$ is called the *statical deflection*, produced by the force $f(t)$. We get this expression from (86) by neglecting the term x'', i.e. neglecting the dynamic nature of the action of the force.

The second term is the correction that has to be given to the statical effect in order to obtain the actual dynamic effect of the force. This second term can be put in the form:

$$-\frac{1}{k^2}\int_0^t f'(u)\cos k(t-u)\,\mathrm{d}u =$$

$$= -\frac{1}{k^2}\cos kt\int_0^t f'(u)\cos ku\,\mathrm{d}u - \frac{1}{k^2}\sin kt\int_0^t f'(u)\sin ku\,\mathrm{d}u. \tag{90}$$

Let us consider the interval of increase of the force, so that we have $t < T_1$. In order to simplify the argument, we shall assume that the first derivative $f'(t)$, which is positive in the interval $(0, T_1)$, is diminishing, i.e. that the growth of the force becomes slower in the course of time. We show that, given this assumption, the two integrals on the right-hand side of equation (90) are small in absolute value. We shall only discuss the integral containing $\sin ku$, since the other integral can be treated in a like manner.

We subdivide the total interval of integration $(0, t)$ into half-periods of the free vibration, $\tau/2 = \pi/k$, and let the number of full half-periods included in t be m, so that:

$$m\frac{\tau}{2} < t < (m+1)\frac{\tau}{2}.$$

We now have

$$\int_0^t f'(u)\sin ku\,\mathrm{d}u = \int_0^{\frac{\tau}{2}} f'(u)\sin ku\,\mathrm{d}u + \int_{\frac{\tau}{2}}^{\tau} f'(u)\sin ku\,\mathrm{d}u + \ldots +$$

$$+\int_{(m-1)\frac{\tau}{2}}^{m\frac{\tau}{2}} f'(u)\sin ku\,\mathrm{d}u + \int_{m\frac{\tau}{2}}^{t} f'(u)\sin ku\,\mathrm{d}u,$$

and the last interval $(m\tau/2, t)$ will in general be less than $\tau/2$.

Since $\sin ku$ does not change sign in each of the sub-intervals into which the total interval has been divided, we can apply the mean value theorem [I, 95] and, bearing in mind that $k\tau = 2\pi$, we can write

$$\int_{s\frac{\tau}{2}}^{(s+1)\frac{\tau}{2}} f'(u)\sin ku\,\mathrm{d}u = f'(u_s)\int_{s\frac{\tau}{2}}^{(s+1)\frac{\tau}{2}} \sin ku\,\mathrm{d}u =$$

$$= \frac{1}{k} f'(u_s)\left[\cos ku\right]_{u=s\frac{\tau}{2}}^{u=(s+1)\frac{\tau}{2}} =$$

$$= -\frac{1}{k}f'(u_s)\left[\cos(s+1)\pi - \cos s\pi\right] = (-1)^s\frac{2}{k}f'(u_s) = (-1)^s\frac{\tau}{\pi}f'(u_s),$$

where

$$s\frac{\tau}{2} < u_s < (s+1)\frac{\tau}{2} \; (s = 0,\ 1,\ 2, \ldots m-1).$$

Similarly, we have for the last interval:

$$\int_{m\frac{\tau}{2}}^{t} f'(u) \sin ku \, du = (-1)^m \frac{\tau}{\pi} \theta f'(u_m), \text{ where } 0 < \theta < 1 \text{ and } m\frac{\tau}{2} < u_m < t.$$

Hence we have:

$$\int_0^t f'(u) \sin ku \, du = \frac{\tau}{\pi} [f'(u_0) - f'(u_1) + f'(u_2) - \ldots +$$
$$+ (-1)^{m-1} f'(u_{m-1}) + (-1)^m \theta f'(u_m)].$$

In view of our assumption regarding $f'(t)$, the terms of the alternating series decrease in absolute value on moving away from the initial term; the total sum therefore has the $(+)$ sign but is less than the first term [I, 123]:

$$0 < \int_0^t f'(u) \sin ku \, du < \frac{1}{\pi} \tau f'(u_1).$$

For τ small, $\tau f'(u_1)$ is approximately equal to the increment of $f(u)$ in the interval $(u_1, u_1 + \tau)$ [I, 50], i.e. $\tau f'(u_1)$ is roughly equal to the change in the force in an interval of time equal to the period of free vibration.

If this interval is so small by comparison with the total interval of increase of the force that the above change in the force can be reckoned negligible, the integrals

$$\int_0^t f'(u) \sin ku \, du \text{ and } \int_0^t f'(u) \cos ku \, du$$

will be small in absolute value, and the second term on the right-hand side of equation (89) will, by (90), be a small quantity compared with the first term. Precisely the same argument applies for the interval in which the force is decreasing. Hence, *if the period of free vibration is small compared with the total duration of the action of the force, the deviation produced by the force can be found from the statical deviation.*

It follows from the above discussion that τ must be so small by comparison with T that the change in the force during the interval τ can be neglected.

If the derivative $f'(t)$ is not always decreasing during the interval of increase of the force, but has a single maximum, as is often the case in practice, the argument remains the same in essence as that given above. The only difference lies in the fact that, in summing the alternating series, it has to be divided into two parts, and the prevailing term in the sum will be a middle term, corresponding to the particular interval in which the maximum of $f'(t)$ occurs.

The possibility of determining statically the deflection due to the external force is of importance in devices designed for recording this force. We shall

take as an example the indicator of a steam engine. This amounts to a cylinder with a close-fitting piston. The latter is subject to the pressure of the steam and compresses an elastic spring.

Let s be the area of the piston, $f_1(t)$ the pressure of the steam, k_1^2 the rigidity of the spring, m the mass of the piston, and x its displacement. The equation of motion of the piston is

$$mx'' = -k_1^2 x + sf_1(t), \quad \text{or} \quad x'' + k^2 x = f(t),$$

where

$$k^2 = \frac{k_1^2}{m} \quad \text{and} \quad f(t) = \frac{sf_1(t)}{m} .$$

The value of x is given by (89). The second term on the right-hand side of this expression represents the instrument error. For the error to be small, the period of free vibration of the piston on the spring must be small compared with the duration of the action of the force. Given this, the reading of the indicator will approximate closely to the curve of $f(t)$, i.e. to the curve of the external force (to within a constant factor). If the pressure increases so rapidly, however, that the change in pressure is significant during an interval equal to the period of free vibration, the indicator readings will diverge considerably from the pressure curve.†

36. The strength of a thin elastic rod, compressed by longitudinal forces (Euler's problem). If a thin, straight elastic rod AB, the ends of which can move along the line AB (Fig. 26), is subjected to two forces P, acting on its ends and compressing it along its axis, distortion of the axis of the rod, leading to its collapse, can occur at a critical value of the force. The problem of finding the force capable of producing such distortion (the problem of the so-called "longitudinal bending" of the rod) was first stated and solved by Euler.

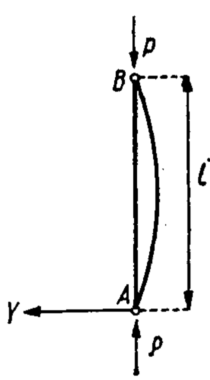

FIG. 26

Let l be the length of the rod AB, E the modulus of elasticity of the material of the rod, and I the moment of inertia of its cross-section, which we can take as constant over all its length [16].

Let OX be taken from the end A along the axis of the rod to the end B, and let y denote the ordinate of the elastic curve of the rod. The differential equation of the elastic curve becomes in this case:††

$$EI \frac{d^2y}{dx^2} = -Py \tag{91}$$

or, putting $q^2 = P/EI$:

$$\frac{d^2y}{dx^2} + q^2 y = 0. \tag{92}$$

† A more detailed account of this problem may be found in A. N. Krilov's article, "Nekotorie zamechaniya o kresherah i indikatorah" in Izvestiya Akademii Nauk, 1909.

†† The bending moment of one of the forces P is evidently $(-Py)$ for any section of the rod.

The general solution of this equation is:

$$y = C_1 \cos qx + C_2 \sin qx. \tag{93}$$

The fact that the ends A, B must remain on axis OX gives us the conditions:

$$y|_{x=0} = y|_{x=l} = 0. \tag{94}$$

We notice that these are not initial conditions. Initial conditions specify the value of the function y and of its derivative y' for a definite value of x. Conditions (94) specify only y, though for two values of the independent variable, at the ends of the interval $(0, l)$; they are called, in fact, *boundary* conditions.

We substitute $x = 0$ and $x = l$ in the general solution (93):

$$0 = C_1; \quad 0 = C_1 \cos ql + C_2 \sin ql \quad \text{and} \quad C_1 = 0; \quad C_2 \sin ql = 0. \tag{95}$$

These equations have the obvious solution $C_1 = C_2 = 0$, which. by (93), gives $y = 0$, i.e. the straight form of the rod. For distortion of the axis to be possible, we must have $C_2 \neq 0$, which means that $\sin ql = 0$. Hence q must take one of the values:

$$q = \frac{\pi s}{l} \; (s = 0, \; 1, \; 2, \ldots). \tag{96}$$

The first solution $s = 0$ makes q and y zero, and again gives the straight elastic curve. The least non-zero value of q is obtained for $s = 1$:

$$q_1 = \frac{\pi}{l} \,.$$

On substituting this value in the equation $q^2 = P/EI$, we get the *least value of the force capable of producing distortion:*

$$P_1 = EIq_1^2 = \frac{\pi^2 EI}{l^2}, \tag{97}$$

or the so-called *critical force* (Euler's formula).

The curve along which the rod bends for $P = P_1$ will have the equation:

$$y = C_2 \sin \frac{\pi}{l} x.$$

i.e. consists of a half sine wave (Fig. 26). The state of equilibrium is unstable, and considerable deformations are possible.

We find, on setting $s = 2$ in (96):

$$q_2 = \frac{2\pi}{l} \,.$$

The equation of the axis of bending of the rod now becomes:

$$y = C_2 \sin \frac{2\pi}{l} x,$$

and the bending curve consists here of two half waves.

The force P_2 needed to produce this deformation is:

$$P_2 = EIq_2^2 = \frac{4\pi^2 EI}{l^2},$$

and is thus four times greater than the previous force.

On giving successive integral values to s, we get all the possible equilibrium forms of the bending axis of the rod. These will consist of a corresponding number of half sine waves, whilst the forces required for the appearance of these distortions will be proportional to the square of the number of half waves.

It may be pointed out that differential equation (91) is approximate, in the sense that the curvature of the bending axis of the rod is taken equal to the second derivative; the equation therefore only applies in regard to small deformations of the rod. The conclusions drawn from general solution (93) of the equation are not justified as regards forces P which produce considerable bendings of the rod, and can clearly lead to absurd results.

Numerous experiments with long, thin rods have shown that the rod at first preserves its straight shape with gradually increasing P, then suffers considerable distortion of its axis on P reaching a value near P_1, as defined by (97); the bending thereafter increases with great rapidity as P continues to increase.

The role of the boundary conditions (94) must be mentioned. Given the initial conditions, the solution of a linear equation is uniquely defined. A different situation arises with boundary conditions, as we have seen. Particular values (96) of the coefficient q in equation (92) exist, such that, given boundary conditions (94), the equation has, apart from the obvious solution $y = 0$, solutions which are defined up to an arbitrary constant factor. We shall meet with the same situation in the example below [37].

37. Rotating shaft. Experiment shows that the following effect occurs on rapid rotation of a long, thin shaft: as the angular velocity increases, it reaches a certain value $\omega = \omega_1$ at which the shaft no longer remains straight but begins to wobble; as ω increases further, stability is again achieved for a time, then is lost again at $\omega = \omega_2$, and so on. We explain the reason for this effect and the method of calculating the critical velocities: $\omega_1, \omega_2, \ldots$

Generally speaking, a rotating shaft has a straight shape at equilibrium, but at the above-mentioned critical velocities the shaft can have a bent dynamical equilibrium shape in addition to the straight equilibrium shape; with this, any chance factor can lead to distortion of the shaft, which causes it to wobble.

Let the shaft be supported at its ends $x = 0$ and $x = l$, and let y denote the amount of bending, as usual. Each element dx of the bent rotating shaft is subjected to a centrifugal force $(p/g)\,\omega^2 y\,dx$, where p is the weight per unit length of the shaft and g is the acceleration due to gravity. On taking this force as a continuously distributed load, we obtain from equations (25) and (32) [16]:

$$EI\frac{d^4 y}{dx^4} = \frac{p\omega^2}{g}\,y,$$

or, writing

$$q = \sqrt[4]{\frac{p\omega^2}{gEI}}, \tag{98}$$

we have:

$$y^{(\mathrm{iv})} - q^4 y = 0. \tag{99}$$

The corresponding characteristic equation $r^4 - q^4 = 0$ has roots: $\pm q$, $\pm qi$, and the general solution of (99) becomes:

$$y = C_1 \mathrm{e}^{qx} + C_2 \mathrm{e}^{-qx} + C_3 \cos qx + C_4 \sin qx.$$

The deflection and bending moment must both be zero at the supported ends; we thus have the four boundary conditions:

$$y\Big|_{x=0} = \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\Big|_{x=0} = y\Big|_{x=l} = \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\Big|_{x=l} = 0,$$

These can easily be seen to reduce to the system of equations:

$$\left.\begin{array}{l} C_1 \mathrm{e}^{ql} + C_2 \mathrm{e}^{-ql} - C_3 \cos ql - C_4 \sin ql = 0; \\ C_1 \mathrm{e}^{ql} + C_2 \mathrm{e}^{-ql} + C_3 \cos ql + C_4 \sin ql = 0; \\ C_1 + C_2 + C_3 = 0; \qquad C_1 + C_2 - C_3 = 0. \end{array}\right\} \tag{100}$$

The solution

$$C_1 = C_2 = C_3 = C_4 = 0 \tag{101}$$

corresponds to the obvious identity $y = 0$, i.e. to the straight shape of the shaft. We now find the values of q for which system (100) has solutions differing from (101).

The first two equations give:

$$C_1 = -C_2; \qquad C_3 = 0.$$

We get by substitution in the last two equations:

$$C_1 = C_2 = C_3 = 0; \qquad C_4 \sin ql = 0.$$

If $C_4 \neq 0$, we must have $\sin ql = 0$, which gives the values for q:

$$q = \frac{s\pi}{l} \; (s = 1, 2, \ldots). \tag{102}$$

On using (98), we get the following expression for the critical velocities:

$$\omega_s = \frac{s^2 \pi^2}{l^2} \sqrt{\frac{EIg}{p}} \; (s = 1, 2, 3, \ldots).$$

38. Symbolic method. We now come to a fresh method of integrating a single linear equation or a set of linear equations with constant coefficients. The method can be applied, with suitable generalization, to more complex problems. It consists essentially in a symbolic notation for the operation of differentiation with respect to the

independent variable t, by writing a D to the left of the function to be differentiated. Thus, if x is a function of t,

$$Dx = \frac{\mathrm{d}x}{\mathrm{d}t},$$

and in general, for any positive integral s:

$$D^s x = \frac{\mathrm{d}^s x}{\mathrm{d}t^s}. \tag{103}$$

If a is a constant, obviously

$$D^s (ax) = aD^s x, \tag{104}$$

i.e. the law of transposition holds for the product of a symbolic factor and a constant. If $F(D)$ is a polynomial in D with constant coefficients:

$$F(D) = a_0 D^n + a_1 D^{n-1} + \ldots + a_{n-1} D + a_n,$$

the operation $F(D)\,x$ is defined as:

$$F(D)\,x = a_0 D^n x + a_1 D^{n-1} x + \ldots + a_{n-1} Dx + a_n x =$$
$$= a_0 \frac{\mathrm{d}^n x}{\mathrm{d}t^n} + a_1 \frac{\mathrm{d}^{n-1} x}{\mathrm{d}t^{n-1}} + \ldots + a_{n-1} \frac{\mathrm{d}x}{\mathrm{d}t} + a_n x.$$

If $\varphi(D)$ is the product of two polynomials $\varphi_1(D)$ and $\varphi_2(D)$, we have, using (104) and the obvious equality $D^{n_1} (D^{n_2} x) = D^{n_1+n_2} x$:

$$\varphi_1 (D) [\varphi_2 (D)\, x] = \varphi(D)\, x,$$

where the factors $\varphi_1(D)$ and $\varphi_2(D)$ can be transposed.

Evidently, in the same way,

$$[\varphi_1 (D) + \varphi_2 (D)]\, x = \varphi_1 (D)\, x + \varphi_2 (D)\, x,$$

and the result obtained is independent of the order of the terms $\varphi_1(D)$ and $\varphi_2(D)$.

The ordinary rules of addition, subtraction and multiplication thus extend to the symbolic polynomials now introduced.

By (104), a constant factor can be taken outside the sign of a symbolic polynomial, so that, along with (104), we have

$$F(D)\,(ax) = aF(D)\,x;$$

of course, this is not permissible with factors dependent on t. We now prove the formula:

$$F(D)\,(\mathrm{e}^{mt} x) = \mathrm{e}^{mt} F(D + m)\, x, \tag{105}$$

where m is a constant. The expression indicates that *a factor of the form* e^{mt} *can be taken outside the sign of a symbolic polynomial after substituting* $(D + m)$ *for* D *in the latter.*

The expression $F(D)(e^{mt}x)$ consists of terms of the type $a_{n-s}D^s(e^{mt}x)$, and it is sufficient to prove (105) for all such terms, i.e. we need only prove

$$D^s(e^{mt}x) = e^{mt}(D+m)^s x. \tag{106}$$

If we use Leibniz's formula for differentiating a product, we can write [I, 53]:

$$D^s(e^{mt}x) = \frac{d^s(e^{mt}x)}{dt^s} = (e^{mt})^{(s)}x + C_s^1(e^{mt})^{(s-1)}x' + \\ + C_s^2(e^{mt})^{(s-2)}x'' + \ldots + C_s^k(e^{mt})^{(s-k)}x^{(k)} + \ldots e^{mt}x^{(s)},$$

where the superscript in brackets indicates the order of the derivative with respect to t, and C_s^k is the number of combinations of k from s elements. Since $(e^{mt})^{(p)} = m^p e^{mt}$ and $x^{(p)} = D^p x$, we can write, on taking e^{mt} outside the brackets:

$$D^s(e^{mt}x) = e^{mt}(m^s x + C_s^1 m^{s-1}Dx + C_s^2 m^{s-2}D^2x + \ldots + \\ + C_s^k m^{s-k}D^k x + \ldots + D^s x) = e^{mt}(m^s + C_s^1 m^{s-1}D + \\ + C_s^2 m^{s-2}D^2 + \ldots + C_s^k m^{s-k}D^k + \ldots + D^s)x.$$

But the right-hand side is identical with the right-hand side of (106); (106) is thus proved, which amounts to proving (105).

We now define negative powers of D as operations the inverse of differentiation, i.e. we define $D^{-s}f(t)$ as the solution of the equation

$$D^s x = f(t), \tag{107}$$

where, in order to give a precise meaning to the symbol $D^{-s}f(t)$, we agree to take the solution which satisfies the zero initial conditions:

$$x|_{t=t_0} = x'|_{t=t_0} = \ldots = x^{(s-1)}|_{t=t_0} = 0. \tag{108}$$

In other words, we shall take [15]

$$D^{-s}f(t) = \frac{1}{(s-1)!}\int_{t_0}^{t}(t-u)^{s-1}f(u)\,du. \tag{109}$$

The general solution of equation (107) now becomes [15]:

$$x = D^{-s} f(t) + P_{s-1}(t) = \frac{1}{(s-1)!} \int_{t_0}^{t} (t-u)^{s-1} f(u)\, \mathrm{d}u + P_{s-1}(t), \tag{110}$$

where $P_{s-1}(t)$ is a polynomial in t of degree $(s-1)$ with arbitrary coefficients.

We define the more general operation $(D-a)^{-s} f(t)$ as the solution of the equation

$$(D-a)^s x = f(t), \tag{111}$$

satisfying conditions (108). To find this solution, we bring in a new unknown z instead of x, where:

$$x = \mathrm{e}^{at} z. \tag{112}$$

On substituting in (111) and using the rule expressed by (105), we obtain the equation for z:

$$\mathrm{e}^{at}(D + a - a)^s z = f(t) \quad \text{or} \quad D^s z = \mathrm{e}^{-at} f(t). \tag{113}$$

The solution of this equation, which satisfies the conditions:

$$z|_{t=t_0} = z'|_{t=t_0} = \ldots = z^{(s-1)}|_{t=t_0} = 0, \tag{114}$$

can be determined in accordance with (109), provided we write $\mathrm{e}^{-at} f(t)$ here in place of $f(t)$:

$$z = \frac{1}{(s-1)!} \int_{t_0}^{t} (t-u)^{s-1} \mathrm{e}^{-au} f(u)\, \mathrm{d}u.$$

But it follows from the formulae:

$$D^j x = D^j \mathrm{e}^{at} z = \mathrm{e}^{at}(D+a)^j z \; (j = 0,\ 1,\ 2, \ldots, s-1)$$

that, if z satisfies conditions (114), x, as defined by (112), satisfies conditions (108). On substituting the expression found for z in (112), we get the required solution of (111):

$$(D-a)^{-s} f(t) = \frac{\mathrm{e}^{at}}{(s-1)!} \int_{t_0}^{t} (t-u)^{s-1} \mathrm{e}^{-au} f(u)\, \mathrm{d}u. \tag{115}$$

The general solution of this equation is obtained if we multiply the general solution of equation (113) by e^{at}, i.e. the general solution will be:

$$x = (D-a)^{-s} f(t) + e^{at} P_{s-1}(t) = \\ = \frac{e^{at}}{(s-1)!}\int_{t_0}^{t}(t-u)^{s-1} e^{-au} f(u)\,du + e^{at} P_{s-1}(t), \tag{116}$$

where $P_{s-1}(t)$ is a polynomial in t of degree $(s-1)$ with arbitrary coefficients.

In particular, setting $f(t) = 0$, we get the general solution of the equation

$$(D-a)^s x = 0 \tag{117}$$

as

$$x = e^{at} P_{s-1}(t). \tag{118}$$

39. Linear homogeneous equations of higher orders with constant coefficients. A linear homogeneous equation of the nth order with constant coefficients is of the form:

$$x^{(n)} + a_1 x^{(n-1)} + \ldots + a_{n-1} x' + a_n x = 0. \tag{119}$$

If we denote differentiation with respect to t by the symbolic operator D and introduce the polynomial:

$$\varphi(D) = D^n + a_1 D^{n-1} + \ldots + a_{n-1} D + a_n,$$

we can write the equation as:

$$\varphi(D)\, x = 0. \tag{120}$$

The characteristic equation corresponding to equation (119) is:

$$r^n + a_1 r^{n-1} + \ldots + a_{n-1} r + a_n = 0, \tag{121}$$

with, say, roots $r_1, r_2, \ldots, r_m$ of multiplicities $k_1, k_2, \ldots, k_m$, where

$$k_1 + k_2 + \ldots + k_m = n\,. \tag{122}$$

On factorizing the polynomial $\varphi(D)$, we can write equation (120) in the form:

$$(D-r_1)^{k_1}(D-r_2)^{k_2}\ldots(D-r_m)^{k_m} x = 0. \tag{123}$$

The equation

$$(D-r_m)^{k_m} x = 0, \tag{124}$$

has, by (118) [38], the general solution

$$x = e^{r_m t} P_{k_m-1}(t), \tag{125}$$

where $P_{k_m-1}(t)$ is a polynomial of degree $(k_m - 1)$ with arbitrary coefficients.

Function (125) will clearly also be a solution of equation (123). We see this by substituting (125) in (123), when we get zero as a result of the operation $(D - r_m)^{k_m}$; the operations

$$(D - r_1)^{k_1} (D - r_2)^{k_2} \dots (D - r_{m-1})^{k_{m-1}},$$

multiplied by zero, evidently also give zero. We could now transpose the factors so that some other factor, say $(D - r_s)^{k_s}$, stood next to x. It may be seen, by means of this device, that a series of particular solutions exists:

$$x_s = e^{r_s t} P_{k_s-1}(t) \qquad (s = 1, 2, \dots, m), \tag{126}$$

where $P_{k_s-1}(t)$ is a polynomial of degree $(k_s - 1)$ with arbitrary coefficients.

By assigning to s all the values from 1 to m in (126), and adding all the solutions thus obtained, we arrive at the solution of equation (123) [26]:

$$x = e^{r_1 t} P_{k_1-1}(t) + e^{r_2 t} P_{k_2-1}(t) + \dots + e^{r_m t} P_{k_m-1}(t). \tag{127}$$

Each polynomial $P_{k_s-1}(t)$ of degree $(k_s - 1)$ with arbitrary coefficients contains altogether k_s arbitrary constants, and therefore, by relationship (122), solution (127) contains in all n arbitrary constants. In view of this, it may be surmised that (127) represents the general solution of equation (119), i.e. that every solution of this equation is included in (127).

This was proved above by expression (118) of [38] for the case of $m = 1$, so that it remains to show that, if our assertion is justified for the case of $(m - 1)$ factors of the form $(D - r_s)^{k_s}$, it is also justified for m factors. The proof is as follows. Equation (123) can be rewritten as:

$$(D - r_1)^{k_1} (D - r_2)^{k_2} \dots (D - r_{m-1})^{k_{m-1}} y = 0,$$

where

$$y = (D - r_m)^{k_m} x.$$

We suppose that our statement is proved for $(m - 1)$ factors, so that we have the general solution for y:

$$y = (D - r_m)^{k_m} x = e^{r_1 t} Q_{k_1-1}(t) + e^{r_2 t} Q_{k_2-1}(t) + \dots + \\ + e^{r_{m-1} t} Q_{k_{m-1}-1}(t),$$

where the $Q_{k_s-1}(t)$ are arbitrary polynomials of degree $(k_s - 1)$. On writing

$$x = e^{r_m t} z, \tag{128}$$

taking $e^{r_m t}$ outside the sign of the symbolic polynomial, and dividing both sides of the equation by $e^{r_m t}$, we get:

$$D^{k_m} z = e^{(r_1 - r_m)t} Q_{k_1-1}(t) + e^{(r_2 - r_m)t} Q_{k_2-1}(t) + \ldots + \\ + e^{(r_{m-1} - r_m)t} Q_{k_{m-1}-1}(t).$$

We get the general expression for z on integrating the right-hand side k_m times with respect to t and adding a polynomial of degree $(k_m - 1)$ [15]. Now we know [I, 201] that the integral of the product of the exponential function e^{at} and a polynomial of degree k in t is of the same form. Thus, z must be of the form:

$$z = e^{(r_1 - r_m)t} P_{k_1-1}(t) + e^{(r_2 - r_m)t} P_{k_2-1}(t) + \ldots + \\ + e^{(r_{m-1} - r_m)t} P_{k_{m-1}-1}(t) + P_{k_m-1}(t).$$

On using (128), we see that x must be as given by (127), which is what we required to prove.

In particular, if the roots of the characteristic equation are all simple, all the $P_{k_s-1}(t)$ are polynomials of zero degree ($k_s = 1$), in other words, they are arbitrary constants C_s; here, the general solution of the equation has the form:

$$x = C_1 e^{r_1 t} + C_2 e^{r_2 t} + \ldots + C_n e^{r_n t}.$$

Assuming that the coefficients of equation (121) are real, some of its roots may nevertheless be complex. The terms in solution (127) corresponding to these complex roots are easily reducible to real form by passing from exponential to trigonometric functions. Suppose that (121) has a pair of conjugate complex roots $(\gamma \pm \delta i)$ of multiplicity k. The solution corresponding to these will be of the form

$$e^{(\gamma+\delta i)t} S_{k-1}(t) + e^{(\gamma-\delta i)t} T_{k-1}(t) = e^{\gamma t}[e^{\delta t i} S_{k-1}(t) + e^{-\delta t i} T_{k-1}(t)],$$

where $S_{k-1}(t)$ and $T_{k-1}(t)$ are polynomials of degree $(k-1)$ with arbitrary coefficients. If we write

$$e^{\delta t i} = \cos \delta t + i \sin \delta t; \qquad e^{-\delta t i} = \cos \delta t - i \sin \delta t,$$

we obtain the solution in the form

$$e^{\gamma t}[U_{k-1}(t) \cos \delta t + V_{k-1}(t) \sin \delta t],$$

where $U_{k-1}(t)$ and $V_{k-1}(t)$ are polynomials of degree $(k-1)$ with arbitrary coefficients, related to $S_{k-1}(t)$ and $T_{k-1}(t)$ via the expressions

$$U_{k-1}(t) = S_{k-1}(t) + T_{k-1}(t); \quad V_{k-1}(t) = i[S_{k-1}(t) - T_{k-1}(t)].$$

The above discussion leads us to the following rule [27]: *equation* (119) *must be integrated by first forming the corresponding characteristic equation* (121) *and finding its roots. For every real solution $r = r'$ of multiplicity k', there is a corresponding solution of the form:*

$$e^{r't} P_{k'-1}(t),$$

where $P_{k'-1}(t)$ is a polynomial of degree $(k'-1)$ with arbitrary coefficients; and for every pair of conjugate complex roots $r = (\gamma \pm \delta i)$ of multiplicity k, there is a corresponding solution of the form

$$e^{\gamma t}[U_{k-1}(t)\cos\delta t + V_{k-1}(t)\sin\delta t],$$

where $U_{k-1}(t)$ and $V_{k-1}(t)$ are polynomials of degree $(k-1)$ with arbitrary coefficients. Addition of all the solutions thus obtained gives us the general solution of equation (119). *The polynomials reduce to arbitrary constants in the case of simple roots.*

40. Linear non-homogeneous equations with constant coefficients. A linear non-homogeneous equation is of the form

$$\varphi(D)x = f(t), \tag{129}$$

where $f(t)$ is a given function. Its general solution is obtained [25] by adding to the general solution of the corresponding homogeneous equation, which we know already how to obtain, a particular solution of it, which we must now proceed to find. We use the symbolic method and split the rational fraction $1/\varphi(D)$ into partial fractions [I, 196]:

$$\frac{1}{\varphi(D)} = \sum_{s=1}^{m}\sum_{q=1}^{k_s}\frac{A_s^{(q)}}{(D-r_s)^q}.$$

Let the function $\xi(t)$ be defined by

$$\xi(t) = \sum_{s=1}^{m}\sum_{q=1}^{k_s}\frac{A_s^{(q)}}{(D-r_s)^q}f(t), \tag{130}$$

the meaning of this expression being fully defined since, in accordance with (115) of [38], each member of the right-hand side has a definite meaning:

$$\frac{A_s^{(q)}}{(D-r_s)^q} f(t) = A_s^{(q)} \frac{e^{r_s t}}{(q-1)!} \int_{t_0}^{t} (t-u)^{q-1} e^{-r_s u} f(u)\, du. \tag{131}$$

It is easily shown that (130) in fact gives a solution of (129). We consider

$$\varphi(D)\,\xi(t) = \sum_{s=1}^{m} \sum_{q=1}^{k_s} \varphi(D) \frac{A_s^{(q)}}{(D-r_s)^q} f(t).$$

By definition of the symbol $(D - r_s)^{-q}$, if we operate on the right-hand side of (131) with $(D - r_s)^q$, the result is $A_s^{(q)} f(t)$. Since the polynomial $\varphi(D)$ is divisible by $(D - r_s)^q$, we have $\varphi(D) = \varphi_{sq}(D)$ $(D - r_s)^q$, where $\varphi_{sq}(D)$ is a polynomial. Hence we can re-write the previous formula as:

$$\varphi(D)\,\xi(t) = \sum_{s=1}^{m} \sum_{q=1}^{k_s} A_s^{(q)} \varphi_{sq}(D) f(t).$$

But it immediately follows from the expansion of $1/\varphi(D)$ that

$$\sum_{s=1}^{m} \sum_{q=1}^{k_s} A_s^{(q)} \varphi_{sq}(D) = 1, \text{ and therefore } \varphi(D)\,\xi(t) = f(t),$$

so that a solution of equation (129) is actually given by (130). We see from this that finding a solution of (129) for any given function $f(t)$ reduces to splitting a rational fraction into partial fractions and then integrating.

The particular solution of equation (129) is found more simply in some particular cases by the method of undetermined coefficients, as demonstrated in [29], than by the general formula (130).

We remark that formulae (71) and (72) of [32] are easily derived by using the above symbolic method.

41. Example. We take as an example the equation

$$x^{(\mathrm{iv})} + 2x'' + x = t \cos t. \tag{132}$$

Here, the characteristic equation becomes

$$r^4 + 2r^2 + 1 = 0 \quad \text{or} \quad (r^2 + 1)^2 = 0. \tag{133}$$

This has a pair of conjugate, double roots $r = \pm i$. The general solution of the homogeneous equation corresponding to (132) is

$$(C_1 t + C_2) \cos t + (C_3 t + C_4) \sin t. \tag{134}$$

We see by comparing the right-hand side of the given equation with (43) of [29] that here, $k = 0$, $l = 1$, and $P(x) = 1$, $Q(x) = 0$. The numbers $k + li = \pm i$ coincide with the pair of double roots, so that the solution of (132) must, by [29], be sought in the form

$$x = t^2 \left[(at + b) \cos t + (ct + d) \sin t\right]. \tag{135}$$

The problem will be simplified by writing the right-hand side of (132) in the exponential form. If we do this, and at the same time write the left-hand side in the symbolic form, (132) becomes:

$$(D^2 + 1)^2 x = \frac{t}{2} e^{it} + \frac{t}{2} e^{-it}. \tag{136}$$

We shall have to look for the solution in the form

$$x = t^2 (at + b) e^{it} + t^2 (ct + d) e^{-it}. \tag{137}$$

We substitute this expression in the left-hand side of the equation:

$$(D + i)^2 (D - i)^2 t^2 (at + b) e^{it} + (D + i)^2 (D - i)^2 t^2 (ct + d) e^{-it} = \frac{t}{2} e^{it} + \frac{t}{2} e^{-it}.$$

On taking e^{it} and e^{-it} outside the symbolic polynomials in accordance with rule (111), we get:

$$e^{it} (D + 2i)^2 D^2 (at^3 + bt^2) + e^{-it} (D - 2i)^2 D^2 (ct^3 + dt^2) = \frac{t}{2} e^{it} + \frac{t}{2} e^{it},$$

or, on writing the second derivative instead of D^2:

$$e^{it} (D^2 + 4iD - 4) (6at + 2b) + e^{-it} (D^2 - 4iD - 4) (6ct + 2d) = \frac{t}{2} e^{it} + \frac{t}{2} e^{-it}.$$

We carry out the differentiations:

$$[-24at + (24ai - 8b)] e^{it} + [-24ct - (24ci + 8d)] e^{-it} = \frac{t}{2} e^{it} + \frac{t}{2} e^{-it}.$$

Hence, by the method of undetermined coefficients:

$$-24a = \frac{1}{2}; \quad 24ai - 8b = 0; \quad -24c = \frac{1}{2}; \quad 24ci + 8d = 0,$$

or

$$a = -\frac{1}{48}; \quad b = \frac{1}{16} i; \quad c = -\frac{1}{48}; \quad d = \frac{1}{16} i.$$

On substituting in (137), we obtain the solution:

$$x = -\frac{t^3}{24} \cos t - \frac{t^2}{8} \sin t, \tag{138}$$

and the general solution of equation (132) becomes:

$$x = (C_1 t + C_2) \cos t + (C_3 t + C_4) \sin t - \frac{t^3}{24} \cos t - \frac{t^2}{8} \sin t. \tag{139}$$

42. Euler's equation. This equation has the form

$$t^n x^{(n)} + a_1 t^{n-1} x^{(n-1)} + \ldots + a_{n-1} t x' + a_n x = 0, \qquad (140)$$

where $a_1, a_2, \ldots, a_n$ are constants. We show that it reduces to an equation with constant coefficients on replacing the independent variable t by τ, defined by the expression

$$t = \mathrm{e}^{\tau}. \qquad (141)$$

We denote differentiation with respect to t by the symbolic factor D as previously, whilst symbolic δ will denote differentiation with respect to τ. We obviously have

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = \frac{\mathrm{d}x}{\mathrm{d}t} \cdot \frac{\mathrm{d}t}{\mathrm{d}\tau} = \mathrm{e}^{\tau} \frac{\mathrm{d}x}{\mathrm{d}t},$$

or, in symbolic notation:

$$Dx = \mathrm{e}^{-\tau} \delta x. \qquad (142)$$

If we operate on the left-hand side with D, and on the right-hand side with the equivalent $\mathrm{e}^{-\tau} \delta$, we get:

$$D^2 x = \mathrm{e}^{-\tau}\delta(\mathrm{e}^{-\tau} \delta)\, x.$$

On taking the factor $\mathrm{e}^{-\tau}$ outside the δ sign, in accordance with the rule of (111), we find:

$$D^2 x = \mathrm{e}^{-2\tau} (\delta - 1)\, \delta x = \mathrm{e}^{-2\tau} \delta\, (\delta - 1)\, x.$$

This expression, together with (142), suggests the following general formula:

$$D^s x = \mathrm{e}^{-s\tau} \delta(\delta - 1) \ldots (\delta - s + 1)\, x. \qquad (143)$$

We have to show that, if the formula is true for s symbolic factors, it is true for $(s + 1)$ factors. Let (143) be assumed true, and let us operate on the left-hand side with D, and on the right-hand side with the equivalent $\mathrm{e}^{-\tau} \delta$; we get:

$$D^{s+1} x = \mathrm{e}^{-\tau} \delta\, [\mathrm{e}^{-s\tau} \delta\, (\delta - 1) \ldots (\delta - s + 1)\, x],$$

whence, taking $\mathrm{e}^{-s\tau}$ outside the δ sign:

$$D^{s+1} x = \mathrm{e}^{-\tau} \mathrm{e}^{-s\tau} (\delta - s)\, \delta\, (\delta - 1) \ldots (\delta - s + 1)\, x =$$

$$= \mathrm{e}^{-(s+1)\tau} \delta\, (\delta - 1) \ldots (\delta - s + 1)\, (\delta - s)\, x,$$

which shows that (143) is true for any s.

On writing t for e^{τ}, the formula becomes:

$$t^s D^s x = \delta\, (\delta - 1) \ldots (\delta - s + 1)\, x. \qquad (144)$$

It follows that, as a result of transformation (141), every term $a_{n-s} t^s x^{(s)}$ on the left-hand side of (140) is replaced by a term

$$a_{n-s}\,\delta(\delta-1)\ldots(\delta-s+1)\,x,$$

which does not contain the independent variable τ; hence we obtain the linear equation with constant coefficients:

$$[\delta(\delta-1)\ldots(\delta-n+1)+a_1\delta(\delta-1)\ldots(\delta-n+2)+\ldots+ \\ +a_{n-1}\delta+a_n]\,x=0. \tag{145}$$

The characteristic equation corresponding to this will be:

$$r(r-1)\ldots(r-n+1)+a_1 r(r-1)\ldots(r-n+2)+\ldots+ \\ +a_{n-1}r+a_n=0, \tag{146}$$

so that the general solution of (145) is:

$$x=e^{r_1\tau}P_{k_1-1}(\tau)+e^{r_2\tau}P_{k_2-1}(\tau)+\ldots+e^{r_m\tau}P_{k_m-1}(\tau),$$

where the r_s are roots of (146) of multiplicity k_s and the $P_{k_s-1}(\tau)$ are polynomials of degree (k_s-1) with arbitrary coefficients.

On returning to the original variable via (141), we get the solution of equation (140) as:

$$x=t^{r_1}P_{k_1-1}(\log t)+t^{r_2}P_{k_2-1}(\log t)+\ldots+t^{r_m}P_{k_m-1}(\log t). \tag{147}$$

If the roots of equation (146) are all simple, the solution of (140) becomes:

$$x=C_1t^{r_1}+C_2t^{r_2}+\ldots+C_nt^{r}n. \tag{148}$$

Equation (146) is obtained, as is easily seen, if the solution of (140) is sought in the form $x=t^r$.

Given a non-homogeneous equation of the form

$$t^n x^{(n)}+a_1t^{n-1}x^{(n-1)}+\ldots+a_{n-1}tx'+a_nx=t^aP(\log t), \tag{149}$$

where $P(\log t)$ is a polynomial of degree p in $\log t$, it may easily be seen, by using transformation (141), that the solution can be sought in the form:

$$x=(\log t)^s\,t^a\,Q(\log t), \tag{150}$$

where $Q(\log t)$ is a polynomial of degree p in $\log t$ and s is the number of roots of equation (146) equal to a.

We can take, instead of (140), the more general equation of the type:

$$(ct+d)^n x^{(n)}+a_1(ct+d)^{n-1}x^{(n-1)}+\ldots+ \\ +a_{n-1}(ct+d)\,x'+a_nx=0. \tag{151}$$

The variables must here be transformed by the formula:

$$ct + d = e^{\tau},$$

instead of by (141), whilst instead of (144) we now have the expression:

$$(ct + d)^s D^s x = c^s \delta(\delta - 1) \ldots (\delta - s + 1)x,$$

with the aid of which equation (151) also reduces to an equation with constant coefficients.

43. Systems of linear equations with constant coefficients. The position of a mechanical system is in many cases defined, not by one, but by several independent quantities $q_1, q_2, \ldots, q_k$, which are referred to as parametric coordinates. The number k of these gives *the number of degrees of freedom.* In the case of rotation of a rigid body about a fixed axis, for instance, we have one degree of freedom, this being the angle θ by which the body has turned about the axis. The rotation of a body about a fixed point provides three degrees of freedom, and the parametric coordinates can be taken, for instance, as the Eulerian angles, φ, ψ and θ, which are familiar from the dynamics of rigid bodies. The motion of a point-mass on a plane, a sphere, or any other surface, has two degrees of freedom. The parametric coordinates may be taken as the ordinary rectangular coordinates x and y in the case of a plane, or as the longitude φ and latitude ψ in the case of a sphere.

With the motion of a mechanical system, its parametric coordinates $q_1, q_2, \ldots, q_k$ are functions of time t, and are defined by a system of differential equations with initial conditions. In particular, when considering the small oscillations of a system about a position of equilibrium, for which the corresponding parametric values are

$$q_1 = q_2 = \ldots = q_k = 0,$$

usually only first order terms in q_s and dq_s/dt are retained in the differential equations, so that we have a linear system with constant coefficients. Every equation will in general contain all the q_s and their first and second order derivatives with respect to t.

Given two degrees of freedom, the system will have the form:

$$\left.\begin{aligned} a_1 q_1'' + b_1 q_1' + c_1 q_1 + a_2 q_2'' + b_2 q_2' + c_2 q_2 = 0; \\ d_1 q_1'' + e_1 q_1' + f_1 q_1 + d_2 q_2'' + e_2 q_2' + f_2 q_2 = 0, \end{aligned}\right\} \qquad (152)$$

where q_1', q_1'', q_2', q_2'' are derivatives of q_1 and q_2 with respect to t.

On using the above notation of symbolic D for differentiation with respect to t, we can write (152) in a different form:

$$\left.\begin{aligned}(a_1 D^2 + b_1 D + c_1)\,q_1 + (a_2 D^2 + b_2 D + c_2)\,q_2 = 0;\\ (d_1 D^2 + e_1 D + f_1)\,q_1 + (d_2 D^2 + e_2 D + f_2)\,q_2 = 0.\end{aligned}\right\} \quad (153)$$

If external forces act on the system, there will be known functions of t instead of zero on the right-hand sides of the equations.

The initial conditions have the form:

$$q_1\,|_{t=0} = q_{10}; \quad q_1'\,|_{t=0} = q_{10}'; \quad q_2\,|_{t=0} = q_{20}; \quad q_2'\,|_{t=0} = q_{20}',$$

where q_{10}, q_{10}', q_{20}, q_{20}' are given numbers, and the general solution of system (153) must contain four arbitrary constants.

We show how integration of system (153) reduces to integration of a single linear equation of the fourth order with one unknown [20]. We do this by bringing in an auxiliary function V of t, putting:

$$q_1 = -(a_2 D^2 + b_2 D + c_2)\,V; \quad q_2 = (a_1 D^2 + b_1 D + c_1)\,V. \quad (154)$$

On substituting these expressions for q_1 and q_2 in equations (153), we see that the first equation will be satisfied for any V, so that it remains to choose V such that the second equation will also be satisfied.

Substitution of expressions (154) in this second equation gives us a fourth order equation for V:†

$$\begin{aligned}&[(a_1 D^2 + b_1 D + c_1)(d_2 D^2 + e_2 D + f_2) -\\ &- (a_2 D^2 + b_2 D + c_2)(d_1 D^2 + e_1 D + f_1)]\,V = 0.\end{aligned} \quad (155)$$

Having found V, we get q_1 and q_2 from (154) by straightforward differentiation.

Let r_1, r_2, r_3, r_4 be the non-repeated roots of the characteristic equation:

$$\begin{aligned}&(a_1 r^2 + b_1 r + c_1)(d_2 r^2 + e_2 r + f_2) -\\ &- (a_2 r^2 + b_2 r + c_2)(d_1 r^2 + e_1 r + f_1) = 0,\end{aligned} \quad (156)$$

so that

$$V = C_1 e^{r_1 t} + C_2 e^{r_2 t} + C_3 e^{r_3 t} + C_4 e^{r_4 t}. \quad (157)$$

† We assume that $a_1 d_2 - a_2 d_1 \neq 0$; this is always the case when considering the motion of a material system.

We substitute this expression in (154), recall that $De^{rt} = re^{rt}$ and $D^2 e^{rt} = r^2 e^{rt}$, and get a general expression for q_1 and q_2. It will consist of a linear combination of four solutions, each of which contains an arbitrary constant factor. The solution $V = C_1 e^{r_1 t}$ gives, for instance:

$$q_1 = -C_1 (a_2 r_1^2 + b_2 r_1 + c_2) e^{r_1 t}; \quad q_2 = C_1 (a_1 r_1^2 + b_1 r_1 + c_1) e^{r_1 t}. \tag{158}$$

If equation (156) has complex roots, as is usual in applications the solution of equation (155) can usefully be written in trigonometric form, so that the solutions for V:

$$C_1 e^{at} \cos bt \quad \text{and} \quad C_2 e^{at} \sin bt,$$

correspond to a pair of conjugate roots $r = a \pm bi$.

Similarly, if (156) has a double root $r_1 = r_2$, the solutions become

$$C_1 e^{r_1 t} \quad \text{and} \quad C_2 t e^{r_1 t}.$$

We now notice the case when the above method does not lead to the general solution for q_1 and q_2, containing four arbitrary constants. Let equation (156) become, for a certain root r_1:

$$a_1 r_1^2 + b_1 r_1 + c_1 = a_2 r_1^2 + b_2 r_1 + c_2 = 0. \tag{159}$$

Expressions (158) for q_1 and q_2 now vanish identically, and the general solution of the system will not contain the arbitrary constant C_1. We can try to restore the lost constant by using the equations

$$q_1 = (d_2 D^2 + e_2 D + f_2) V; \quad q_2 = -(d_1 D^2 + e_1 D + f_1) V. \tag{160}$$

instead of (154), when we introduce the auxiliary function V.

With this, the second of equations (153) will be satisfied for any V, whilst substitution of expressions (160) in the first equation (153) will give us the same equation (155) as above for V. The root r_1 of the characteristic equation (156) now gives, instead of (158), the expressions for q_1 and q_2:

$$q_1 = C_1 (d_2 r_1^2 + e_2 r_1 + f_2) e^{r_1 t}; \quad q_2 = -C_1 (d_1 r_1^2 + e_1 r_1 + f_1) e^{r_1 t}.$$

Provided one at least of the factors $(d_1 r_1^2 + e_1 r_1 + f_1)$ and $(d_2 r_1^2 + e_2 r_1 + f_2)$ does not vanish, the solution corresponding to the root $r = r_1$ of (156) has now been restored.

The case still remains to be considered when, in addition to relationships (159), we have

$$d_1 r_1^2 + e_1 r_1 + f_1 = d_2 r_1^2 + e_2 r_1 + f_2 = 0. \tag{161}$$

With this, the method given above does not lead to restoration of the solution corresponding to root $r = r_1$ of equation (156). But since (159) and (161) are valid, all the quadratic expressions in brackets on the left-hand side of (156) have the root $r = r_1$, i.e. are divisible by $(r - r_1)$. It follows that $r = r_1$ must be a repeated root of (156). We confine ourselves to the case when $r = r_1$ is a double root, and indicate the corresponding solutions of the system. These two solutions will be:

$$q_1 = C_1 \mathrm{e}^{r_1 t}; \quad q_2 = 0 \tag{162}$$

$$q_1 = 0; \qquad q_2 = C_2 \mathrm{e}^{r_1 t}. \tag{163}$$

If, in fact, we substitute say expressions (162) in the left-hand side of either of equations (153), we obtain an identity, by (159) and (161).

These solutions are distinct, since q_2 is identically zero in the first, whereas it differs from zero in the second.

We remark that if, in the case of a repeated root $r_1 = r_2$, say, one of relationships (159) is not fulfilled, we obtain, on substituting

$$V = C_1 \mathrm{e}^{r_1 t} \text{ and } V = C_2 t\mathrm{e}^{r_1 t}$$

in expressions (154), the solution (158) and a solution which contains t as a factor:

$$q_1 = -C_2 (a_2 r_1^2 + b_2 r_1 + c_2)\, t\mathrm{e}^{r_1 t} + C_2 p_1 \mathrm{e}^{r_1 t};$$
$$q_2 = C_2 (a_1 r_1^2 + b_1 r_1 + c_1)\, t\mathrm{e}^{r_1 t} + C_2 p_2 \mathrm{e}^{r_1 t},$$

where p_1 and p_2 are definite constants.

The general solution of the non-homogeneous system:

$$\left.\begin{aligned} (a_1 D^2 + b_1 D + c_1) q_1 + (a_2 D^2 + b_2 D + c_2) q_2 = f_1(t); \\ (d_1 D^2 + e_1 D + f_1) q_1 + (d_2 D^2 + e_2 D + f_2) q_2 = f_2(t), \end{aligned}\right\} \tag{164}$$

consists, as in the case of a single equation, in the sum of the general solution of the corresponding homogeneous system (153) and any particular solution of the non-homogeneous system. If the free terms $f_1(t)$ and $f_2(t)$ are of the form

$$A_0 \mathrm{e}^{\alpha t} \cos \beta t + B_0 \mathrm{e}^{\alpha t} \sin \beta t = D\mathrm{e}^{\alpha t} \sin(\beta t + \varphi),$$

the particular solutions can be sought in the form

$$q_1 = A_1 \mathrm{e}^{\alpha t} \cos \beta t + B_1 \mathrm{e}^{\alpha t} \sin \beta t; \quad q_2 = A_2 \mathrm{e}^{\alpha t} \cos \beta t + B_2 \mathrm{e}^{\alpha t} \sin \beta t,$$

provided only that $(\alpha \pm \beta i)$ is not a root of equation (156). On substituting this expression in the left-hand side of equation (164) and equating coefficients of $e^{\alpha t} \cos \beta t$ and $e^{\alpha t} \sin \beta t$, we get equations for determining A_1, B_1, A_2, B_2.

The particular solutions of system (164) can be obtained for any $f_1(t)$ and $f_2(t)$ in the same way as in the case of a single equation [40]. On solving system (164) for q_1 and q_2, we get, for example, for q_1:

$$q_1 = \frac{d_2 D^2 + e_2 D + f_2}{\Delta(D)} f_1(t) - \frac{a_2 D^2 + b_2 D + c_2}{\Delta(D)} f_2(t),$$

where, for brevity, $\Delta(D)$ denotes the symbolic polynomial on the left-hand side of equation (155). On expanding the rational fractions and using the value given in [38] for the symbolic factor $(D - r)^{-k}$, we obtain the required solution of system (164).

We further remark that, by using the arguments of [20], we can easily reduce the integration of a system of linear equations with constant coefficients to the integration of a single linear equation with constant coefficients. We give a general method in Volume III for integrating a system of equations with constant coefficients.

44. Examples. 1. We consider the system:

$$\frac{d^2 y}{dx^2} = z + x; \quad \frac{d^2 z}{dx^2} = y + 2x,$$

where y and z are required functions of x. On finding z from the first equation:

$$z = \frac{d^2 y}{dx^2} - x$$

and substituting in the second, we get the fourth order equation for y:

$$\frac{d^4 y}{dx^4} - y = 2x, \tag{165}$$

the general solution of which is found by the usual rule as

$$y = C_1 e^x + C_2 e^{-x} + C_3 \cos x + C_4 \sin x - 2x.$$

We substitute this expression in (165) and get the expression for z:

$$z = C_1 e^x + C_2 e^{-x} - C_3 \cos x - C_4 \sin x - x.$$

2. We take the system of three first order equations:

$$\frac{dx}{dt} = y + z; \quad \frac{dy}{dt} = z + x; \quad \frac{dz}{dt} = x + y, \tag{166}$$

where x, y and z are required as functions of t. We solve the first equation with respect to y:

$$y = \frac{dx}{dt} - z \tag{167}$$

and substitute the expression obtained in the remaining two equations, which gives:

$$\frac{d^2x}{dt^2} - \frac{dz}{dt} = z + x; \qquad \frac{dz}{dt} = x + \frac{dx}{dt} - z. \tag{168}$$

On substituting the expression for dz/dt given by the second equation into the first, we get a second order equation containing only the one unknown x (exceptional case):

$$\frac{d^2x}{dt^2} - \frac{dx}{dt} - 2x = 0,$$

the general solution of which is:

$$x = C_1e^{2t} + C_2e^{-t}. \tag{169_1}$$

We substitute this in the second of equations (168) and get a first order equation for z:

$$\frac{dz}{dt} + z = 3C_1e^{2t},$$

the general solution of which is:

$$z = C_3e^{-t} + C_1e^{2t}. \tag{169_2}$$

On substituting expressions (169_1) and (169_2) in (167), we obtain for y:

$$y = C_1e^{2t} - (C_2 + C_3)\,e^{-t}. \tag{169_3}$$

We are presented here with the exceptional case referred to in (20). Instead of obtaining a single differential equation of the third order, we have obtained an equation of the second order and a further equation of the first order.

3. We encounter systems of linear equations with constant coefficients in the study of electrical oscillations, as well as when considering small oscillations of mechanical systems about the equilibrium position. Let two circuits be linked magnetically, i.e. the current in one circuit produces a magnetic field which induces an electromotive force in the second circuit. If i_1, i_2 are the currents in the two circuits, the induced e. m. f. in the first circuit will be $M\,di_2/dt$, and in the second, $M\,di_1/dt$, where M is the constant mutual inductance. If we assume that there is no current source in either circuit, we have the equations:

$$L_1\frac{d^2i_1}{dt^2} + R_1\frac{di_1}{dt} + \frac{1}{C_1}i_1 + M\frac{d^2i_2}{dt^2} = 0; \tag{170}$$

$$M\frac{d^2i_1}{dt^2} + L_2\frac{d^2i_2}{dt^2} + R\frac{di_2}{dt} + \frac{1}{C_2}i_2 = 0, \tag{171}$$

where L_1, R_1, C_1 are the self-inductance, resistance, and capacity in the first circuit, and L_2, R_2, C_2 are the corresponding values for the second circuit.

We use this example to show how one of the unknown functions can be eliminated and a single fourth order equation obtained with one unknown, without introducing an auxiliary function V.

We substitute for $\mathrm{d}^2 i_2/\mathrm{d}t^2$ from (171) into (170), and get the equation

$$(L_1L_2 - M^2)\frac{\mathrm{d}^2 i_1}{\mathrm{d}t^2} + L_2R_1\frac{\mathrm{d}i_1}{\mathrm{d}t} + \frac{L_2}{C_1}i_1 - R_2M\frac{\mathrm{d}i_2}{\mathrm{d}t} - \frac{M}{C_2}i_2 = 0. \qquad (172)$$

On differentiating this equation, and substituting the expression for $M\mathrm{d}^2i_2/\mathrm{d}t^2$:

$$M\frac{\mathrm{d}^2 i_2}{\mathrm{d}t^2} = -L_1\frac{\mathrm{d}^2 i_1}{\mathrm{d}t^2} - R_1\frac{\mathrm{d}i_1}{\mathrm{d}t} - \frac{1}{C_1}i_1 \qquad (173)$$

got from (170), we obtain:

$$(L_1L_2 - M^2)\frac{\mathrm{d}^3 i_1}{\mathrm{d}t^3} + (L_1R_2 + L_2R_1)\frac{\mathrm{d}^2 i_1}{\mathrm{d}t^2} + \left(\frac{L_2}{C_1} + R_1R_2\right)\frac{\mathrm{d}i_1}{\mathrm{d}t} +$$

$$+ \frac{R_2}{C_1}i_1 + \frac{M}{C_2}\frac{\mathrm{d}i_2}{\mathrm{d}t} = 0. \qquad (174)$$

After differentiating once more, and again substituting for $M\mathrm{d}^2i_2/\mathrm{d}t^2$ from (173), we finally arrive at a linear equation of the fourth order for i_1:

$$(L_1L_2 - M^2)\frac{\mathrm{d}^4 i_1}{\mathrm{d}t^4} + (L_1R_2 + L_2R_1)\frac{\mathrm{d}^3 i_1}{\mathrm{d}t^3} + \left(\frac{L_1}{C_2} + \frac{L_2}{C_1} + R_1R_2\right)\frac{\mathrm{d}^2 i_1}{\mathrm{d}t^2} +$$

$$+ \left(\frac{R_1}{C_2} + \frac{R_2}{C_1}\right)\frac{\mathrm{d}i_1}{\mathrm{d}t} + \frac{1}{C_1C_2}i_1 = 0. \qquad (175)$$

If we had eliminated i_1 instead of i_2, we should have got precisely the same fourth order equation for i_2. Its corresponding characteristic equation is

$$(1 - k^2)r^4 + 2(g_1 + g_2)r^3 + (n_1^2 + n_2^2 + 4g_1g_2)r^2 +$$

$$+ 2(g_1n_2^2 + g_2n_1^2)r + n_1^2n_2^2 = 0, \qquad (176)$$

where we have written for brevity:

$$k = \frac{M}{\sqrt{L_1L_2}}; \qquad n_1 = \frac{1}{\sqrt{L_1C_1}}; \qquad n_2 = \frac{1}{\sqrt{L_2C_2}}; \qquad g_1 = \frac{R_1}{2L_1}; \qquad g_2 = \frac{R_2}{2L_2}$$

Equation (176) can also be written in the form:

$$(r^2 + 2g_1r + n_1^2)(r^2 + 2g_2r + n_2^2) - k^2r^4 = 0. \qquad (177)$$

If there were no magnetic coupling, we should have to put $M = 0$ in equations (170) and (171), and we should obtain two separate equations, defining the discharge processes in the circuits:

$$\frac{\mathrm{d}^2 i_1}{\mathrm{d}t^2} + 2g_1\frac{\mathrm{d}i_1}{\mathrm{d}t} + n_1^2 i_1 = 0 \quad \text{and} \quad \frac{\mathrm{d}^2 i_2}{\mathrm{d}t^2} + 2g_2\frac{\mathrm{d}i_2}{\mathrm{d}t} + n_2^2 i_2 = 0. \qquad (178)$$

Both circuits are usually oscillatory, in other words, the characteristic equations corresponding to differential equations (178):

$$r^2 + 2g_1r + n_1^2 = 0 \quad \text{and} \quad r^2 + 2g_2r + n_2^2 = 0, \qquad (179)$$

have complex roots, i.e. $g_1^2 - n_1^2 < 0$ and $g_2^2 - n_2^2 < 0$, or

$$\frac{R_1}{2L_1} < \frac{1}{\sqrt{L_1 C_1}} \text{ and } \frac{R_2}{2L_2} < \frac{1}{\sqrt{L_2 C_2}},$$

or alternatively

$$\frac{R_1}{2} < \sqrt{\frac{L_1}{C_1}} \text{ and } \frac{R_2}{2} < \sqrt{\frac{L_2}{C_2}},$$

With $k = 0$, equation (177) gives two pairs of conjugate complex roots [the roots of equations (179)]. With the small values of M, such as are usually found in practice, (177) again has two pairs of conjugate complex roots, with negative real parts: $r_{1,2} = -a \pm bi$ and $r_{3,4} = -c \pm di$; so that the general expression for i_1 becomes:

$$i_1 = C_1 e^{-at} \cos bt + C_2 e^{-at} \sin bt + C_3 e^{-ct} \cos dt + C_4 e^{-ct} \sin dt.$$

We remark that, on knowing i_1, we can obtain i_2 also without further integration. All we do is find di_2/dt from equation (174), substitute the expression obtained in equation (172), and thus get a linear equation in i_2. The expression for i_2 will contain terms of the same form as in i_1, with coefficients that are linear combinations of the constants C_1, C_2, C_3 and C_4.

If we neglect the resistances, i.e. take $g_1 = g_2 = 0$, and in addition, assume that the circuits are tuned to the same frequency, i.e. $n_1 = n_2 = n$, equation (177) becomes

$$(1 - k^2) r^4 + 2n^2 r^2 + n^4 = 0,$$

whence

$$r^2 = \frac{-n^2 \pm kn^2}{1 - k^2} = -\frac{n^2}{1 \pm k},$$

and

$$r_1, r_2 = \pm \frac{n}{\sqrt{1+k}} i; \quad r_3, r_4 = \pm \frac{n}{\sqrt{1-k}} i \qquad (i = \sqrt{-1}).$$

The solution corresponding to these purely imaginary roots is of trigonometric form. It follows that, given magnetic coupling between two circuits tuned to the same frequency, two oscillations arise, whose frequencies depend on the common frequency n of the circuits and on the constant k which characterises the magnetic coupling, in accordance with the relationships:

$$n' = \frac{n}{\sqrt{1+k}}; \qquad n'' = \frac{n}{\sqrt{1-k}}.$$

§ 4. Integration with the aid of power series

45. Integration of a linear equation, using a power series. The solutions of a linear equation of higher order than the first with variable coefficients are not generally expressible in terms of elementary

functions, as we mentioned earlier, and the integration of such an equation does not in general reduce to a quadrature. The most useful method is to represent the required solution as a power series, as already mentioned in [13]. This device is particularly applicable to linear differential equations. We shall confine ourselves to the second order equation:

$$y'' + p(x)y' + q(x)y = 0. \tag{1}$$

Let the coefficients $p(x)$ and $q(x)$ be given in the form of power series, expanded in positive integral powers of x, so that the equation has the form:

$$y'' + (a_0 + a_1 x + a_2 x^2 + \ldots)y' + (b_0 + b_1 x + b_2 x^2 + \ldots)y = 0. \tag{2}$$

We notice that the coefficient of y'' is taken equal to unity.

We seek the solution of (2) as a further power series:

$$y = \sum_{s=0}^{\infty} \alpha_s x^s. \tag{3}$$

On substituting for y and its derivatives from (3) into (2), we get:

$$\sum_{s=2}^{\infty} s(s-1)\alpha_s x^{s-2} + \sum_{s=0}^{\infty} a_s x^s \sum_{s=1}^{\infty} s\alpha_s x^{s-1} + \sum_{s=0}^{\infty} b_s x^s \sum_{s=0}^{\infty} \alpha_s x^s = 0.$$

By cross-multiplying the power series, collecting like terms and equating to zero the coefficients of the various powers of x on the left-hand side of the equation written, we get a series of equations:

$$\left.\begin{array}{l|l}
x^0 & 2\cdot 1\alpha_2 + a_0\alpha_1 + b_0\alpha_0 = 0 \\
x^1 & 3\cdot 2\alpha_3 + 2a_0\alpha_2 + a_1\alpha_1 + b_0\alpha_1 + b_1\alpha_0 = 0 \\
x^2 & 4\cdot 3\alpha_4 + 3a_0\alpha_3 + 2a_1\alpha_2 + a_2\alpha_1 + b_0\alpha_2 + b_1\alpha_1 + b_2\alpha_0 = 0 \\
\cdot\,\cdot & \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot \\
x^s & (s+2)(s+1)\alpha_{s+2} + Q_s(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{s+1}) = 0 \\
\cdot & \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot
\end{array}\right\} \tag{4}$$

Here, $Q_s(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{s+1})$ denotes a homogeneous polynomial of the first degree in its arguments $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{s+1}$.

Each successive equation of the series written contains one more required coefficient than the previous equation. The coefficients α_0 and α_1 remain arbitrary, and play the part of arbitrary constants. The first of equations (4) gives α_2, then the second gives α_3, the third α_4, and so on, so that in general, knowing the preceding $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{s+1}$, we can determine α_{s+2} from the $(s+1)$th equation.

It is convenient to proceed as follows. We find two solutions y_1 and y_2 by the above method, with the values $a_0 = 1$ and $a_1 = 0$ taken for the first solution, and values $a_0 = 0$ and $a_1 = 1$ for the second, this being equivalent to the following initial conditions:

$$y_1 |_{x=0} = 1; \quad y_1' |_{x=0} = 0,$$
$$y_2 |_{x=0} = 0; \quad y_2' |_{x=0} = 1.$$

Every solution of the equation will be a linear combination of these solutions, and if the initial conditions have the form

$$y |_{x=0} = A; \quad y' |_{x=0} = B,$$

obviously,

$$y = Ay_1 + By_2.$$

We have shown above that the coefficients of power series (3) can be determined successively by formal computation. It remains an open question, however, whether the power series thus obtained will be convergent, and whether it will provide a solution of the equation. We give the proof in Volume III of the following proposition: *if the series*

$$p(x) = \sum_{s=0}^{\infty} a_s x^s; \qquad q(x) = \sum_{s=0}^{\infty} b_s x^s$$

are convergent for $|x| < R$, *the power series arrived at by the above method is also convergent for these values of* x, *and gives a solution of equation* (2). *In particular, if* $p(x)$ *and* $q(x)$ *are polynomials in* x, *the power series obtained is convergent for any* x.

Linear equations are often encountered of the form

$$P_0(x) y'' + P_1(x) y' + P_2(x) y = 0, \tag{5}$$

where $P_0(x)$, $P_1(x)$, $P_2(x)$ are polynomials in x. In order to reduce this to form (1), we have to divide both sides by $P_0(x)$, which means taking

$$p(x) = \frac{P_1(x)}{P_0(x)}; \qquad q(x) = \frac{P_2(x)}{P_0(x)}. \tag{6}$$

If the constant term of polynomial $P_0(x)$ differs from zero, i.e. $P_0(0) \neq 0$, $p(x)$ and $q(x)$ can be written as power series, on dividing the polynomials, arranged in increasing powers of x, and the solution of (5) also can be sought as a power series. With this, there is no need to reduce (5) to form (1), the simpler method being to substitute expression (3) directly for y in the left-hand side of (5) then use the method of undetermined coefficients.

We have so far only considered power series, arranged in positive integral powers of x. It is also possible to use series arranged in powers of $(x - a)$.

All that has been said above evidently applies to linear equations of higher order than the second, except that now, instead of the first two coefficients remaining undetermined when finding the solution as a power series, the number of undetermined coefficients becomes equal to the order of the equation.

Given the non-homogeneous linear equation

$$y'' + p(x)\, y' + q(x)\, y = f(x),$$

where both the right-hand side and the coefficients are power series, its particular solution can also be sought as a power series.

We notice one point about expressions (6). Let $P(x)$ and $Q(x)$ be two polynomials in x, where $P(0) \neq 0$. The result of dividing the polynomials can be written as a power series, as mentioned above:

$$\frac{Q(x)}{P(x)} = c_0 + c_1 x + c_2 x^2 + \ldots; \tag{7}$$

but now the questions arise, as to whether the series on the right is convergent, and if so, in what interval; also, is its sum equal to the left-hand side of the equation? The answers to these questions follow quite simply from the theory of functions of a complex variable, which is described in Volume III. Here, we simply give the final result: the power series of (7) is convergent for $|x| < R$, where R is the modulus (or absolute value) of the root of minimum modulus of the equation $P(x) = 0$; furthermore, equation (7) applies for these x. One consequence of this is that, if equation (5) is integrated directly with the aid of power series, the series obtained will in fact be convergent for $|x| < R$, where R is the minimum among the moduli of the roots of the equation $P_0(x) = 0$.

We remark that, if the convergence of series (3) is proved within the interval $(-R, +R)$, it follows directly from this that the sum of the series gives a solution of the equation. To start with, we can find y' and y'' by simple term-by-term differentiation of series (3) [I, 150]. Then, on substituting the expressions for y, y' and y'' in the left-hand side of equation (2), we can multiply term by term the series for y' and y with the series $p(x)$ and $q(x)$, in view of the fact that power series are absolutely convergent [I, 137, 148]. Finally, by the choice of the coefficients a_n from equations (4), we can cancel all the terms on the left-hand side of (2).

46. Examples. 1. We take the equation:

$$y'' - xy = 0.$$

We get by substituting series (3):

$$(2 \cdot 1a_2 + 3 \cdot 2a_3 x + 4 \cdot 3a_4 x^2 + \dots) - x(a_0 + a_1 x + a_2 x^2 + \dots) = 0,$$

whence we find, on equating to zero coefficients of like powers of x:

$$\begin{array}{c|l} x^0 & 2 \cdot 1a_2 = 0 \\ x^1 & 3 \cdot 2a_3 - a_0 = 0 \\ x^2 & 4 \cdot 3a_4 - a_1 = 0 \\ x^3 & 5 \cdot 4a_5 - a_2 = 0 \\ \dots & \dots\dots\dots\dots\dots \\ x^s & (s+2)(s+1)\,a_{s+2} - a_{s-1} = 0 \\ \dots & \dots\dots\dots\dots\dots \end{array}$$

Having put $a_0 = 1$ and $a_1 = 0$, we obtain successively the remaining coefficients:

$$a_2 = 0; \quad a_3 = \frac{1}{2 \cdot 3}; \quad a_4 = a_5 = 0; \quad a_6 = \frac{1}{2 \cdot 3 \cdot 5 \cdot 6}; \quad a_7 = a_8 = 0;$$

$$a_9 = \frac{1}{2 \cdot 3 \cdot 5 \cdot 6 \cdot 8 \cdot 9},$$

i.e. the only coefficients a_s differing from zero are those for which the subscript s is divisible by 3, so that we can write:

$$a_{3k+1} = a_{3k+2} = 0 \text{ and } a_{3k} = \frac{1 \cdot 4 \cdot 7 \dots (3k-2)}{(3k)!}.$$

We have thus obtained the solution

$$y_1 = 1 + \sum_{k=1}^{\infty} \frac{1 \cdot 4 \cdot 7 \dots (3k-2)}{(3k)!} x^{3k}.$$

We get the second solution by putting $a_0 = 0$ and $a_1 = 1$. It is easily shown, by the same method as above, that this second solution is

$$y_2 = x + \sum_{k=1}^{\infty} \frac{2 \cdot 5 \cdot 8 \dots (3k-1)}{(3k+1)!} x^{3k+1}.$$

The power series obtained are convergent for all x.

This may be verified for y_1 by using d'Alembert's test [I, 121]. The ratio of the two successive terms becomes

$$\frac{1 \cdot 4 \cdot 7 \dots (3k+1)}{(3k+3)!} x^{3k+3} : \frac{1 \cdot 4 \cdot 7 \dots (3k-2)}{(3k)!} x^{3k} = \frac{1}{(3k+2)(3k+3)} x^3,$$

and the absolute value of this ratio tends to zero for all x on indefinite increase of k, whence follows the absolute convergence of the series.

2. We take the equation:

$$(1 - x^2)\, y'' - xy' + a^2 y = 0.$$

On substituting series (3) and equating to zero coefficients of x^n, we obtain the following relationships between the a_n:

$$(n+2)(n+1)a_{n+2} - n(n-1)a_n - na_n + a^2 a_n = 0$$

or

$$(n+2)(n+1)a_{n+2} = (n^2 - a^2)a_n.$$

On setting $a_0 = 1$ and $a_1 = 0$, we get the solution:

$$y_1 = 1 - \frac{a^2}{2!}x^2 + \frac{a^2(a^2-4)}{4!}x^4 - \frac{a^2(a^2-4)(a^2-16)}{6!}x^6 + \ldots$$

Similarly, setting $a_0 = 0$ and $a_1 = 1$, we get:

$$y_2 = x - \frac{a^2-1}{3!}x^3 + \frac{(a^2-1)(a^2-9)}{5!}x^5 - \\ - \frac{(a^2-1)(a^2-9)(a^2-25)}{7!}x^7 + \ldots$$

The coefficient $P_0(x)$ in the equation taken, where $P_0(x) = 1 - x^2$, has roots $x = \pm 1$, the absolute values of both these being unity. Hence it follows that the series for y_1 and y_2 must be convergent for $-1 < x < +1$, that is, for $|x| < 1$.

We can verify this with d'Alembert's test. Neglecting the sign, we get for the ratio of two successive terms say of y_1:

$$\frac{a^2(a^2-4)\ldots[a^2-(2n)^2]}{(2n+2)!}x^{2n+2} : \frac{a^2(a^2-4)\ldots[a^2-(2n-2)^2]}{(2n)!}x^{2n} = \\ = \frac{a^2-(2n)^2}{(2n+1)(2n+2)}x^2.$$

We divide numerator and denominator by n^2, which enables us to write the absolute value of the ratio in the form:

$$\left| \frac{4 - \dfrac{a^2}{n^2}}{4 + \dfrac{6}{n} + \dfrac{2}{n^2}} \right| \cdot |x|^2.$$

This ratio tends to $|x|^2$ on indefinite increase of n, and obviously, $|x|^2 < 1$ for $|x| < 1$; thus, by d'Alembert's test, series y_1 is absolutely convergent for $|x| < 1$. It is also clear that the series is divergent for $|x| > 1$, provided a is not an even integer. In this last case, the series breaks off and reduces to a polynomial. Similar conclusions apply regarding series y_2. It can be verified that solutions y_1 and y_2 are expressible in terms of elementary functions, in fact,

$$y_1 = \cos(a \arccos x); \quad y_2 = \frac{1}{a}\sin(a \arccos x).$$

47. Expansion of solutions into generalized power series. Quite a number of equations encountered in applications have the form

$$x^2 y'' + p(x) \cdot xy' + q(x)\, y = 0,$$

where $p(x)$ and $q(x)$ are series, as in equation (2), arranged in positive integral powers of x, or else are polynomials. This equation does not come under type (2), due to the presence of x^2 as a factor of the second derivative. The equation is said to have a *regular singular point* at $x = 0$. We write $p(x)$ and $q(x)$ explicitly as series:

$$x^2 y'' + (a_0 + a_1 x + a_2 x^2 + \ldots)\, xy' + (b_0 + b_1 x + b_2 x^2 + \ldots)\, y = 0, \tag{8}$$

and instead of seeking the solution as the simple power series (3), look for it as the product of some power of x and a power series:

$$y = x^{\varrho} \sum_{s=0}^{\infty}{}' \alpha_s x^s, \tag{9}$$

Here, we can obviously take the first coefficient α_0 as non-zero, in view of the power ϱ not being fixed in the factor outside.

We substitute in the left-hand side of (8) the following expressions for y, y', y'':

$$y = \sum_{s=0}^{\infty}{}' \alpha_s x^{\varrho+s}; \quad y' = \sum_{s=0}^{\infty}{}' (\varrho + s)\, \alpha_s\, x^{\varrho+s-1};$$

$$y'' = \sum_{s=0}^{\infty}{}' (\varrho + s)(\varrho + s - 1)\, \alpha_s x^{\varrho+s-2}.$$

On collecting like terms and equating to zero the coefficients of the various powers of x, we get the series of equations:

$$\left.\begin{array}{l|l}
x^{\varrho} & [\varrho(\varrho-1) + a_0\varrho + b_0]\,\alpha_0 = 0 \\
x^{\varrho+1} & [(\varrho+1)\varrho + a_0(\varrho+1) + b_0]\,\alpha_1 + a_1\varrho\alpha_0 + b_1\alpha_0 = 0 \\
x^{\varrho+2} & [(\varrho+2)(\varrho+1) + a_0(\varrho+2) + b_0]\,\alpha_2 + a_1(\varrho+1)\,\alpha_1 + a_2\varrho\alpha_0\, b_1\alpha_1 + b_2\alpha_0 = 0 \\
\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x^{\varrho+s} & [(\varrho+s)(\varrho+s-s) + a_0(\varrho+s) + b_0]\,\alpha s + Q_s(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{s-1}) = 0. \\
\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array}\right\} \tag{10}$$

The $Q_s(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{s-1})$ denote homogeneous polynomials of the first degree in their arguments $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{s-1}$.

Since $\alpha_0 \neq 0$ by hypothesis, the first of the equations written gives a quadratic equation for ϱ:

$$F(\varrho) = \varrho(\varrho - 1) + a_0\varrho + b_0 = 0. \tag{11}$$

This is called the *indicial equation*.

Let its roots be ϱ_1, ϱ_2. On setting either of these for ϱ in equations (10), we get a series of equations, any given one of which contains one more coefficient α_s than the preceding equation, so that we can find successively $\alpha_1, \alpha_2, \ldots$ The coefficient α_0, which remains arbitrary, plays the role of an arbitrary factor and can be taken say equal to unity.

The first of equations (10) has in fact become an identity after substituting $\varrho = \varrho_1$ or $\varrho = \varrho_2$, whilst the second gives α_1, the third α_2, etc., and in general

the $(s+1)$th gives a_s, assuming $a_0, a_1, \ldots, a_{s-1}$ are already known. The only requirement here is for the coefficients of the a_s in the equations to be non-zero. It is immediately evident that these coefficients can be found from the left-hand side of equation (11) by substituting $(\varrho_1 + s)$ or $(\varrho_2 + s)$ for ϱ, i.e. they are equal to $F(\varrho_1 + s)$ or $F(\varrho_2 + s)$.

Suppose that we have obtained solution (9), starting from the root $\varrho = \varrho_2$ of equation (11). If $F(\varrho_2 + s) \neq 0$ for any positive integer s, the above method can be used successfully for determining the coefficients.

The condition $F(\varrho_2 + s) \neq 0$ is obviously equivalent to the other root ϱ_1 of equation (11) not being of the form $(\varrho_2 + s)$, where s is a positive integer In other words, $(\varrho_1 - \varrho_2)$ must not be a positive integer.

The following conclusions are easily drawn from what has been said.

1. If the roots ϱ_1 and ϱ_2 of equation (11) do not differ by a positive integer or zero, both roots can be used in accordance with the above method to obtain two solutions of the form

$$y_1 = x^{\varrho_1} \sum_{s=0}^{\infty}{}' a_s x^s; \qquad y_2 = x^{\varrho_2} \sum_{s=0}^{\infty}{}' \beta_s x^s \qquad (a_0 \text{ and } \beta_0 \neq 0). \tag{12}$$

2. If $(\varrho_1 - \varrho_2)$ is a positive integer, the above method can be used in general to form only one series:

$$y_1 = x^{\varrho_1} \sum_{s=0}^{\infty} a_s x^s. \tag{13}$$

3. If equation (11) has a repeated root, $\varrho_1 = \varrho_2$, again, only the one series (13) can be formed.

In view of the convergence of the series obtained, a proposition can be stated, similar to that made in [45]: *if the series*

$$\sum_{s=0}^{\infty} a_s x^s \quad \text{and} \quad \sum_{s=0}^{\infty} b_s x^s$$

are convergent for $|x| < R$, *the series formed above are convergent for the same* x *and represent solutions of equation* (8).

The equation

$$x^2 P_0(x)\, y'' + x P_1(x)\, y' + P_2(x)\, y = 0, \tag{14}$$

reduces to the one worked out, where $P_0(x)$, $P_1(x)$, $P_2(x)$ are polynomials or series expanded in positive integral powers of x, and $P_0(0) \neq 0$. Here, as in [45], series (9) can be substituted directly in the left-hand side of (14), without dividing by $P_0(x)$. Furthermore, as in [45], we can take series expanded in positive integral powers of $(x - a)$ instead of x.

In case 1 above, the two solutions (12) are linearly independent, i.e. their ratio is not a constant, as follows at once from the fact that expressions y_1 and y_2 contain different powers of x, x^{ϱ_1} and x^{ϱ_2}, before the summation sign. We have only found one solution (13) in cases 2 and 3. Formula (9) of [24] offers the possiblity of finding a second solution with the aid of a quadrature.

We merely state the result, without proof: if $(\varrho_1 - \varrho_2)$ is a positive integer or zero, we have a solution, in addition to (13), of the form:

$$y_2 = \beta y_1 \log x + x^{\varrho_2} \sum_{s=0}^{\infty} \beta_s x^s. \tag{15}$$

Expression y_2 thus differs in the present case from the ordinary expression (12) in having an additional term of the form $\beta y_1 \log x$. It may happen that the constant β is zero, in which case an expression of the form (12) is obtained for y_2. All these assertions will be proved in Volume III.

48. Bessel's equation. This equation has the form:

$$x^2 y'' + xy' + (x^2 - p^2)\, y = 0, \tag{16}$$

where p is a given constant. It occurs in various problems of astronomy and of pure and applied physics.

Comparison of the equation with (8) shows that here, $a_0 = 1$ and $b_0 = -p^2$; the indicial equation therefore now becomes:

$$\varrho(\varrho - 1) + \varrho - p^2 = 0 \quad \text{or} \quad \varrho^2 - p^2 = 0,$$

with the roots

$$\varrho_1 = p, \quad \varrho_2 = -p.$$

We seek the solution in the form

$$y = x^p (a_0 + a_1 x + a_2 x^2 + \ldots).$$

On substituting in the left-hand side of (16) and equating to zero the coefficients of the powers of x, we get:

$$\begin{array}{c|l} x^{p+1} & \left[(p+1)^2 - p^2\right] a_1 = 0 \\ x^{p+2} & \left[(p+2)^2 - p^2\right] a_2 + a_0 = 0 \\ \cdots & \cdots\cdots\cdots\cdots \\ x^{p+s} & \left[(p+s)^2 - p^2\right] a_s + a_{s-2} = 0. \end{array}$$

We write $a_0 = 1$ then evaluate the coefficients successively, and arrive at the solution:

$$y_1 = x^p \left[1 - \frac{x^2}{2(2p+2)} + \frac{x^4}{2 \cdot 4 \cdot (2p+2)(2p+4)} - \right.$$
$$\left. - \frac{x^6}{2 \cdot 4 \cdot 6 \cdot (2p+2)(2p+4)(2p+6)} + \ldots \right]. \tag{17}$$

A second solution of (16) can be obtained by making use of the second root $\varrho_2 = -p$, and it obviously follows by the simple substitution of $(-p)$ for p in solution (17), since (16) only contains p^2 and remains unchanged on making this substitution:

$$y_2 = x^{-p} \left[1 - \frac{x^2}{2(-2p+2)} + \frac{x^4}{2 \cdot 4 \cdot (-2p+2)(-2p+4)} - \right.$$
$$\left. - \frac{x^6}{2 \cdot 4 \cdot 6(-2p+2)(-2p+4)(-2p+6)} + \ldots \right]. \tag{18}$$

The difference between the roots of the indicial equation is $2p$, so that both the solutions written are satisfactory, provided p is neither an integer nor half an odd integer. Solution (17) gives the Bessel function of the pth order, except for a constant factor; the Bessel function is usually written $J_p(x)$, and is also called a cylindrical function of the first kind. Thus the general solution of equation (16), with p neither an integer nor half an odd integer, is

$$y + C_1 J_p(x) + C_2 J_{-p}(x).$$

The power series appearing in solution (17) is convergent for any x, as may easily be verified by d'Alembert's test.

Now let $p = n$ be an integer (positive). Solution (17) remains in force, whereas solution (18) becomes meaningless, since one of the factors in the denominators of terms of the expansion vanishes as from a certain number. The Bessel function $J_n(x)$ is defined for positive integral $p=n$ with the aid of formula (17) multiplied by the constant factor $1/(2^n \cdot n!)$:

$$J_n(x) = \frac{x^n}{2^n n!}\left[1 - \frac{x^2}{2(2n+2)} + \frac{x^4}{2\cdot 4(2n+2)(2n+4)} - \right.$$
$$\left. - \frac{x^6}{2\cdot 4\cdot 6(2n+2)(2n+4)(2n+6)} + \dots\right]. \qquad (19)$$

The general term in this expansion is

$$(-1)^s \frac{x^{n+2s}}{2^n n!\, 2\cdot 4\cdot 6\cdot \dots 2s\cdot(2n+2)(2n+4)(2n+6)\dots(2n+2s)}.$$

Each of the $2s$ factors which appear in the denominator after $2^n \cdot n!$ contains the factor 2; if we take these out and combine them with the 2^n, we can write the general term in the form:

$$(-1)^s \frac{x^{n+2s}}{2^{n+2s}\cdot n!\,1\cdot 2\cdot 3\dots s\cdot(n+1)(n+2)(n+3)\dots(n+s)} =$$
$$= \frac{(-1)^s}{s!\cdot(n+s)!}\left(\frac{x}{2}\right)^{n+2s},$$

so that formula (19) can be written as

$$J_n(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!\,(n+s)!}\left(\frac{x}{2}\right)^{n+2s}, \qquad (20)$$

where we take $0! = 1$ as usual. In particular, we get for $n = 0$:

$$J_0(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{(s!)^2}\left(\frac{x}{2}\right)^{2s} =$$
$$= \frac{1}{(1!)^2}\left(\frac{x}{2}\right)^2 + \frac{1}{(2!)^2}\left(\frac{x}{2}\right)^4 - \frac{1}{(3!)^2}\left(\frac{x}{2}\right)^6 + \dots \qquad (21)$$

By what was said in [47], equation (16) will have a second solution of the form:

$$K_n(x) = \beta J_n(x)\log x + x^{-n}\sum_{s=0}^{\infty} \beta_s x^s. \qquad (22)$$

in addition to solution (20), when $p = n$ is a positive integer.

This solution obviously tends to infinity for $x = 0$.

The general solution of equation (16) becomes for $p = n$:

$$y = C_1 J_n(x) + C_2 K_n(x). \tag{23}$$

If we want to obtain a solution which is finite at $x = 0$, we have to take constant $C_2 = 0$, i.e. we must confine ourselves to solution (20).

We take a closer look at solution (22) for $p = 0$. Here, the equation becomes:

$$y'' + \frac{1}{x} y' + y = 0, \tag{24}$$

and one of its solutions is given by (21). A second solution can be sought in the form

$$\beta J_0(x) \log x + \beta_0 + \beta_1 x + \beta_2 x^2 + \dots$$

On taking a linear combination of this solution with that already found, we can reduce the free term β_0 to zero, so that the final solution can be sought as

$$\beta J_0(x) \log x + \beta_1 x + \beta_2 x^2 + \dots$$

By substituting this expression in the left-hand side of (24) and using the method of undetermined coefficients, we can successively determine the β_n. We miss out all the working and only give the final expression for the second solution. We set the non-vanishing coefficient β equal to unity here:

$$K_0(x) = J_0(x) \log x + \frac{x^2}{2^2} - \frac{x^4}{2^2 \cdot 4^2}\left(1 + \frac{1}{2}\right) + \frac{x^6}{2^2 \cdot 4^2 \cdot 6^2}\left(1 + \frac{1}{2} + \frac{1}{3}\right) - \dots \tag{25}$$

This is called a *zero order Bessel or cylindrical function of the second kind.*

Finally, let p be half an odd integer, $p = (2n + 1)/2$. Although the difference between the roots of the indicial equation is here the integer $(2n + 1)$, solutions (17) and (18) both remain in force and are linearly independent, inasmuch as we have the factors $x^{(2n+1)/2}$ and $x^{-(2n+1)/2}$ respectively in front of the power series, so that their ratio cannot be constant.

If we substitute $p = 1/2$, for instance, in solution (17), we get the series:

$$x^{1/2}\left[1 - \frac{x^2}{2 \cdot 3} + \frac{x^4}{2 \cdot 4 \cdot 3 \cdot 5} - \frac{x^6}{2 \cdot 4 \cdot 6 \cdot 3 \cdot 5 \cdot 7} + \dots\right] = \frac{1}{\sqrt{x}}\left[x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots\right] = \frac{\sin x}{\sqrt{x}}.$$

On multiplying this series by the constant factor $\sqrt{2/\pi}$, we get the Bessel function $J_{1/2}(x)$:

$$J_{1/2}(x) = \sqrt{\frac{2}{\pi x}} \sin x. \tag{26}$$

Similarly, (18) gives us

$$J_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \cos x, \tag{27}$$

and the general solution of equation (16) for $p = 1/2$ is:

$$y = C_1 J_{1/2}(x) + C_2 J_{-1/2}(x).$$

We note some results without proof: the Bessel function with subscript equal to half an odd integer is in general expressed in terms of elementary functions and has the form:

$$J_{\frac{2n+1}{2}}(x) = \sqrt{\frac{2}{\pi x}}\left[P_n\left(\frac{1}{x}\right)\sin x + Q_n\left(\frac{1}{x}\right)\cos x\right],$$

where $P_n(1/x)$ and $Q_n(1/x)$ are polynomials in $1/x$. In particular:

$$J_{3/2}(x) = \sqrt{\frac{2}{\pi x}}\left(\frac{\sin x}{x} - \cos x\right);$$

$$J_{5/2}(x) = \sqrt{\frac{2}{\pi x}}\left[\left(\frac{3}{x^2} - 1\right)\sin x - \frac{3}{x}\cos x\right];$$

$$J_{-3/2}(x) = \sqrt{\frac{2}{\pi x}}\left(-\sin x - \frac{\cos x}{x}\right);$$

$$J_{-5/2}(x) = \sqrt{\frac{2}{\pi x}}\left[\frac{3}{x}\sin x + \left(\frac{3}{x^2} - 1\right)\cos x\right].$$

Furthermore, we have for any positive integral n:

$$J_{\frac{2n+1}{2}}(x) = (-1)^n \cdot \sqrt{\frac{2}{\pi}} \cdot x^{\frac{2n+1}{2}} \frac{d^n}{d(x^2)^n}\left(\frac{\sin x}{x}\right).$$

The even function $(\sin x)/x$ has to be differentiated n times with respect to x^2 in this formula.

49. Equations reducible to Bessel's equation. We notice some equations that can be reduced to Bessel's equation by a change of variables. Take the equation:

$$x^2 y'' + xy' + (k^2 x^2 - p^2) y = 0, \tag{28}$$

where k is a non-vanishing constant. We take instead of x a new independent variable $\xi = kx$. We now have to substitute as follows in equation (28):

$$y' = \frac{dy}{dx} = \frac{dy}{d\xi} \cdot \frac{d\xi}{dx} = k\frac{dy}{d\xi} \quad \text{and} \quad y'' = \frac{d}{dx}\left(k\frac{dy}{d\xi}\right) = k^2 \frac{d^2 y}{d\xi^2},$$

so that (28) becomes:

$$k^2 x^2 \frac{d^2 y}{d\xi^2} + kx\frac{dy}{d\xi} + (k^2 x^2 - p^2) y = 0$$

or

$$\xi^2 \frac{d^2y}{d\xi^2} + \xi \frac{dy}{d\xi} + (\xi^2 - p^2)\, y = 0,$$

which is Bessel's equation (16) with the independent variable ξ. Using $\xi = kx$, the general solution of equation (28) becomes:

$$y = C_1 J_p(kx) + C_2 J_{-p}(kx), \tag{29}$$

or, if $p = n$ is a positive integer or zero:

$$y = C_1 J_n(kx) + C_2 K_n(kx). \tag{29_1}$$

A broad class of equations reducible to Bessel's equation is revealed by introducing a new independent variable t and a new function u into (16), in accordance with

$$y = t^a u \quad \text{and} \quad x = \gamma t^\beta, \tag{30}$$

where a, β and γ are constants, and β and γ are non-zero. Differentiation gives us at once:

$$\frac{dt}{dx} = \frac{1}{\beta\gamma} t^{1-\beta}; \qquad \frac{dy}{dx} = \frac{1}{\beta\gamma} t^{1-\beta} \frac{dy}{dt}$$

$$\frac{d^2y}{dx^2} = \frac{1}{\beta\gamma} t^{1-\beta} \left(\frac{1}{\beta\gamma} t^{1-\beta} \frac{d^2y}{dt^2} + \frac{1-\beta}{\beta\gamma} t^{-\beta} \frac{dy}{dt} \right),$$

and further,

$$\frac{dy}{dt} = t^a \frac{du}{dt} + a t^{a-1} u; \qquad \frac{d^2y}{dt^2} = t^a \frac{d^2u}{dt^2} +$$

$$+ 2a t^{a-1} \frac{du}{dt} + a(a-1)\, t^{a-2} u.$$

If we substitute the expressions for y, dy/dx, d^2y/dx^2 in equation (16), then replace dy/dt and d^2y/dt^2 by their expressions in terms of u, du/dt, d^2u/dt^2 and make simple rearrangements, we find the equation for u:

$$t^2 \frac{d^2u}{dt^2} + (2a+1)\, t \frac{du}{dt} + (a^2 - \beta^2 p^2 + \beta^2 \gamma^2 t^{2\beta})\, u = 0. \tag{31}$$

Equation (16) has the general solution:

$$y = C_1 J_p(x) + C_2 J_{-p}(x),$$

so that, by (30), the general solution of (31) will be:

$$u = t^{-a} y = C_1 t^{-a} J_p(\gamma t^\beta) + C_2 t^{-a} J_{-p}(\gamma t^\beta), \tag{32}$$

where $J_{-p}(\gamma t_\beta)$ must be replaced by $K_n(\gamma t^\beta)$, if $p = n$ is a positive integer or zero.

Equation (31) has the form:

$$t^2 \frac{d^2u}{dt^2} + a t \frac{du}{dt} + (b + c t^m)\, u = 0, \tag{33}$$

where

$$2a + 1 = a; \quad a^2 - \beta^2 p^2 = b; \quad \beta^2 \gamma^2 = c; \quad 2\beta = m. \tag{34}$$

Given an equation of type (33), with constants c and m not zero, we can conversely find α, β, γ and p from equations (34) and write the general integral in terms of Bessel functions by using (32).

If c or m is zero, (33) is Euler's equation [42] and simply reduces to an equation with constant coefficients.

We consider a particular case of equation (33):

$$t\frac{d^2u}{dt^2} + a\frac{du}{dt} + tu = 0. \tag{35}$$

Multiplication of the equation by t shows that we have here: $b = 0$, $c = 1$, $m = 2$, and a arbitrary. Equations (34) become:

$$2\alpha + 1 = a; \quad \alpha^2 - \beta^2 p^2 = 0; \quad \beta^2\gamma^2 = 1; \quad 2\beta = 2,$$

whence we have

$$\alpha = \frac{a-1}{2}; \quad \beta = 1; \quad \gamma = 1; \quad p = \frac{a-1}{2},$$

and the general integral of (35) becomes, by (32):

$$u = C_1 t^{\frac{1-a}{2}} J_{\frac{a-1}{2}}(t) + C_2 t^{\frac{1-a}{2}} J_{\frac{1-a}{2}}(t),$$

where we have to replace $J_{(1-a)/2}$ by $K_{(a-1)/2}$ if $(1 - a)/2$ happens to be a negative integer or zero. Equation (35) is identical with (24) when $a = 1$.

Equation (33) represents a wide, general class of linear equations that are often encountered in applied mathematics, with general integrals expressible, as we have seen, by means of Bessel functions.

§ 5. Supplementary notes on the theory of differential equations

50. The method of successive approximations for linear equations. Mention has already been made several times of the existence and uniqueness theorem for differential equations. We first prove the theorem for the case of linear equations, and do so by making use of the method of successive approximations, which we introduced originally for the approximate evaluation of the roots of equations [I, 193].

We take for clarity the system of two linear homogeneous equations:

$$\frac{dy}{dx} = p_1(x)\,y + q_1(x)\,z; \quad \frac{dz}{dx} = p_2(x)\,y + q_2(x)\,z \tag{1}$$

with the initial conditions

$$y|_{x=x_0} = y_0; \quad z|_{x=x_0} = z_0. \tag{2}$$

We shall assume that the coefficients in equation (1) are continuous functions of x in a finite, closed interval I ($a \leqslant x \leqslant b$) which contains the initial value x_0, and we take x as belonging to I in future arguments.

The solutions y and z of system (1) must certainly be continuous, differentiable functions, and it is clear from the equations themselves that the derivatives $\mathrm{d}y/\mathrm{d}x$ and $\mathrm{d}z/\mathrm{d}x$ must also be continuous, since the right-hand sides of the equations are continuous with the assumptions made. On integrating the equations term by term from x_0 to x, and taking into account (2), we get:

$$\left.\begin{aligned} y(x) &= y_0 + \int_{x_0}^{x} [p_1(t)\, y(t) + q_1(t)\, z(t)]\, \mathrm{d}t \\ z(x) &= z_0 + \int_{x_0}^{x} [p_2(t)\, y(t) + q_2(t)\, z(t)]\, \mathrm{d}t. \end{aligned}\right\} \tag{3}$$

We have written out the arguments of functions y and z here for ease of working, the variable of integration being denoted by t so as to avoid confusion with the upper limit x. Equations (1) with initial conditions (2) thus bring us to equations (3).

We now prove the converse: if continuous functions $y(x)$ and $z(x)$ satisfy equations (3), they also satisfy equations (1) and initial conditions (2). The latter follows by putting $x = x_0$ in (3) and recalling that an integral vanishes when its limits are the same. Furthermore, differentiation of (3) with respect to x gives us equations (1) [I, 96]. Equations (3) are therefore equivalent, in this sense, to equations (1) with initial conditions (2), and we shall in future consider equations (3) only. We notice that the required functions $y(x)$ and $z(x)$ appear both on the left-hand sides of these equations and under the integral sign on the right-hand sides.

The method of successive approximations works as follows. We take the initial values y_0 and z_0 as first approximations to the required functions y and z and substitute accordingly in the right-hand sides of equations (3); this gives us functions $y_1(x)$ and $z_1(x)$:

$$\left.\begin{aligned} y_1(x) &= y_0 + \int_{x_0}^{x} [p_1(t)\, y_0 + q_1(t)\, z_0]\, \mathrm{d}t \\ z_1(x) &= z_0 + \int_{x_0}^{x} [p_2(t)\, y_0 + q_2(t)\, z_0]\, \mathrm{d}t, \end{aligned}\right\} \tag{4}$$

as second approximations to y and z. These latter functions are clearly continuous in the interval I [I, 96]. We now substitute $y_1(x)$ and $z_1(x)$ for y and z in the right-hand sides of (3) and obtain third approximations $y_2(x)$ and $z_2(x)$:

$$y_2(x) = y_0 + \int_{x_0}^{x} [p_1(t)\,y_1(t) + q_1(t)\,z_1(t)]\,\mathrm{d}t$$

$$z_2(x) = z_0 + \int_{x_0}^{x} [p_2(t)\,y_1(t) + q_2(t)\,z_1(t)]\,\mathrm{d}t,$$

where $y_2(x)$ and $z_2(x)$ are again continuous in the interval I, and so on. In general, the $(n+1)$th approximation will be given by:

$$\left.\begin{aligned} y_n(x) &= y_0 + \int_{x_0}^{x} [p_1(t)\,y_{n-1}(t) + q_1(t)\,z_{n-1}(t)]\,\mathrm{d}t \\ z_n(x) &= z_0 + \int_{x_0}^{x} [p_2(t)\,y_{n-1}(t) + q_2(t)\,z_{n-1}(t)]\,\mathrm{d}t. \end{aligned}\right\} \qquad (5)$$

The coefficients of equations (1) are assumed continuous in I, so that their absolute magnitude will not exceed a definite positive number M in the interval [I, 35]:

$$|p_1(x)| \leqslant M\,; \quad |p_2(x)| \leqslant M\,; \quad |q_2(x)| \leqslant M \quad (x \text{ in } I). \qquad (6)$$

Also, we let m denote the greater of the two positive numbers $|y_0|$ and $|z_0|$, i.e.

$$|y_0| \leqslant m\,; \qquad |z_0| \leqslant m. \qquad (7)$$

We shall in future consider only the part of I to the right of x_0, i.e. we take $x - x_0 \geqslant 0$. The left-hand part can be considered in the same way.

We find the difference between two successive approximations. The first of equations (4) gives:

$$y_1(x) - y_0 = \int_{x_0}^{x} [p_1(t)\,y_0 + q_1(t)\,z_0]\,\mathrm{d}t\,.$$

We replace all the magnitudes under the integral by quantities greater than or equal to their absolute values and find, using (6) and (7) [I, 95]:

$$|y_1(x) - y_0| \leqslant \int_{x_0}^{x} (Mm + Mm)\,\mathrm{d}t\,;$$

i.e.

$$|y_1(x) - y_0| \leqslant m \cdot 2M(x - x_0), \tag{8}$$

whilst similarly

$$|z_1(x) - z_0| \leqslant m \cdot 2M(x - x_0). \tag{8_1}$$

The first of equations (5) becomes, for $n = 2$:

$$y_2(x) = y_0 + \int_{x_0}^{x} [p_1(t)\, y_1(t) + q_1(t)\, z_1(t)]\, \mathrm{d}t,$$

and we have, on subtracting the first of equations (4):

$$|y_2(x) - y_1(x)| = \int_{x_0}^{x} \{p_1(t)[y_1(t) - y_0] + q_1(t)[z_1(t) - (z_0]\}\, \mathrm{d}t.$$

If we again replace the magnitudes under the integral with their absolute values and use (6), (8) and (8_1), we get:

$$|y_2(x) - y_1(x)| \leqslant \int_{x_0}^{x} \{M \cdot m \cdot 2M(t - x_0)\} + \{M \cdot m \cdot 2M(t - x_0)\}\, \mathrm{d}t$$

or

$$|y_2(x) - y_1(x)| \leqslant 2^2 m M^2 \int_{x_0}^{x} (t - x_0)\, \mathrm{d}t = m \cdot 2^2 M^2 \left[\frac{(t - x_0)^2}{2!}\right]_{t=x_0}^{t=x},$$

whence finally,

$$|y_2(x) - y_1(x)| \leqslant m \frac{[2M(x - x_0)]^2}{2!}. \tag{9}$$

Similarly, we have:

$$|z_2(x) - z_1(x)| \leqslant m \frac{[2M(x - x_0)]^2}{2!}. \tag{9_1}$$

We continue by taking the first of equations (5) for $n = 2$ and $n = 3$, and find on subtraction:

$$y_3(x) - y_2(x) = \int_{x_0}^{x} \{p_1(t)[y_2(t) - y_1(t)] + q_1(t)[z_2(t) - z_1(t)]\}\, \mathrm{d}t.$$

We have by using (6), (9) and (9_1) as above:

$$|y_3(x) - y_2(x)| \leqslant \frac{2^3 M^3}{2} \int_{x_0}^{x} (t - x_0)^2\, \mathrm{d}t,$$

whence

$$|y_3(x) - y_2(x)| \leqslant m \frac{[2M(x - x_0)]^3}{3!};$$

and similarly,

$$| z_3(x) - z_2(x) | \leqslant m \frac{[2M(x - x_0)]^3}{3!}.$$

By proceeding thus, we can write the difference between two successive approximations as in general:

$$\left.\begin{aligned} | y_n(x) - y_{n-1}(x) | &\leqslant m \frac{[2M(x - x_0)]^n}{n!} \\ | z_n(x) - z_{n-1}(x) | &\leqslant m \frac{[2M(x - x_0)]^n}{n!}. \end{aligned}\right\} \tag{10}$$

We use these values of the differences to show that $y_n(x)$ and $z_n(x)$ converge uniformly to limits $y(x)$ and $z(x)$ respectively on indefinite increase of n.† We give the proof for the sequence of functions $y_n(x)$. We can replace this sequence by the infinite series

$$y_0 + [y_1(x) - y_0] + [y_2(x) - y_1(x)] + \\ + [y_3(x) - y_2(x)] + \ldots + [y_n(x) - y_{n-1}(x)] + \ldots, \tag{11}$$

the sum of the first $(n + 1)$ terms of which is equal to $y_n(x)$, so that we have to prove the uniform convergence of series (11) [I, 144]. If l is the length of the interval I in which x varies, the first of expressions (10) shows that the terms of (11) do not exceed in absolute value the positive numbers

$$m \frac{(2Ml)^n}{n!} \qquad (n = 1, 2, \ldots),$$

whilst the series consisting of these numbers converges by d'Alembert's test, since the ratio of two adjacent terms is $2Ml/n$, which tends to zero with indefinite increase of n. The same result follows from the expansion of e^x [I, 129]. Series (11) is therefore uniformly convergent in the interval I by Weierstrass's test [I, 147], i.e. $y_n(x)$ tends uniformly in this interval to some function $y(x)$. Similarly, it can be shown that sequence $z_n(x)$ tends uniformly in I to some limit function $z(x)$; thus we have, for x in I:

$$\lim_{n\to\infty} y_n(x) = y(x); \qquad \lim_{n\to\infty} z_n(x) = z(x). \tag{12}$$

Since $y_n(x)$ and $z_n(x)$ are continuous in I, the same can be said for their limits $y(x)$ and $z(x)$ [I, 145].

† It is essential to recall, for what follows, the sections on series with variable terms and uniform convergence in Volume I.

We remark that $(x - x_0)$ must be replaced by $(x_0 - x)$ in the right-hand sides of inequalities (8) and (8_1) when dealing with the part of I to the left of x_0, where $x - x_0 \leqslant 0$. We have to replace $(t - x_0)$ by $(x_0 - t)$ in the further working, and so on. Inequality (10) remains valid throughout I, assuming $(x - x_0)$ is replaced by its absolute value.

We now show that the limit functions satisfy equations (3), i.e. equations (1) with initial conditions (2). This follows immediately from expressions (5), on passing to the limit with $n \to \infty$ on both sides of the equations. Now, $y_n(x)$ and $y_{n-1}(t)$ tend respectively to $y(x)$ and $y(t)$, whilst $z_n(x)$ and $z_{n-1}(t)$ tend to $z(x)$ and $z(t)$, so that we get equations (3) for $y(x)$ and $z(x)$ in the limit. We give a rigorous proof of the passage to the limit. It follows from (12) that:

$$\left.\begin{aligned}\lim_{n\to\infty}\left[p_1(t)\,y_{n-1}(t) + q_1(t)z_{n-1}(t)\right] &= p_1(t)\,y(t) + q_1(t)\,z(t);\\ \lim_{n\to\infty}\left[p_2(t)\,y_{n-1}(t) + q_2(t)\,z_{n-1}(t)\right] &= p_2(t)\,y(t) + q_2(t)\,z(t).\end{aligned}\right\} \qquad (12_1)$$

We show that these passages to the limit occur uniformly with respect to t in the interval I. We confine ourselves to the first expression and work out the difference between the limit and the variable:

$$\begin{aligned}&|\left[p_1(t)\,y(t) + q_1(t)\,z(t)\right] - \left[p_1(t)\,y_{n-1}(t) + q_1(t)\,z_{n-1}(t)\right]| \leqslant\\ &\leqslant |p_1(t)|\,|y(t) - y_{n-1}(t)| + |q_1(t)|\,|z(t) - z(t) - z_{n-1}(t)|.\end{aligned}$$

Since $y_{n-1}(t)$ and $z_{n-1}(t)$ converge uniformly to $y(t)$ and $z(t)$, given any $\varepsilon > 0$, there exists a number N, the same for all t in I, such that

$$|y(t) - y_{n-1}(t)| < \frac{\varepsilon}{2M}; \quad |z(t) - z_{n-1}(t)| < \frac{\varepsilon}{2M} \quad \text{for} \quad n > N.$$

Hence it follows, by (6), that we have the inequality, for any t of I:

$$|[p_1(t)\,y(t) + q_1(t)\,z(t)] - [p_1(t)\,y_{n-1}(t) + q_1(t)\,z_{n-1}(t)]| < \varepsilon \quad \text{for} \quad n > N,$$

which proves the uniform passage to the limit in formulae (12_1) throughout the interval I and in any part of it (x_0, x). We turn back to expressions (5) and use the fact that passage to the limit under the integral sign is possible for uniformly convergent sequences [I, 145]; having passed to the limit, we get equations (3) for $y(x)$ and $z(x)$ from these expressions.

To sum up, we can say that the method of successive approximations has given us a solution of system (1) with initial conditions (2), i.e. we have proved the existence of a solution. We now show that the

solution is unique. Let equations (3) have two solutions: $y(x)$, $z(x)$ and $Y(x)$, $Z(x)$. On substituting in (3) first one solution, then the other and subtracting, we get:

$$\left.\begin{aligned} y(x) - Y(x) &= \int_{x_0}^{x} \{p_1(t)\,[y(t) - Y(t)] + p_1(t)\,[z(t) - Z(t)]\}\,\mathrm{d}t \\ z(x) - Z(x) &= \int_{x_0}^{x} \{[p_2(t)[y(t) - Y(t)] + p_2(t)\,[z(t) - Z(t)]\}\mathrm{d}t. \end{aligned}\right\} \quad (13)$$

We take an interval I_1 of length l_1 to the right of x_0, so that $2Ml_1 = \theta$ is less than unity. We show that the two solutions coincide in this interval. If this were not the case, the absolute values of the differences

$$|y(x) - Y(x)|, \quad |z(x) - Z(x)|$$

would have a positive maximum in I_1, which we denote by the number δ. Let the maximum be attained by the first difference at $x = \xi$, i.e.

$$|y(\xi) - Y(\xi)| = \delta \tag{14}$$

and

$$|y(x) - Y(x)| \leqslant \delta \text{ and } |z(x) - Z(x)| \leqslant \delta \quad (x \text{ in } I_1). \tag{14_1}$$

We take the first of equations (13) for $x = \xi$, and apply the same inequality for the integral as above; this gives us, by (14_1):

$$|y(\xi) - Y(\xi)| < 2M\delta\,(\xi - x_0),$$

whence, using (14) and the fact that ξ belongs to interval I''_1,

$$\delta < 2Ml_1\,\delta, \quad \text{i.e.} \quad \delta < \theta\delta,$$

which is impossible, since $0 < \theta < 1$ by hypothesis.

Our assertion that the solutions y, z and Y, Z do not coincide in the interval I_1 is therefore absurd. We can cover the total interval I by intervals of length l_1, and thus prove the identity of the two solutions throughout I.

We state the final result: *system* (1) *with initial conditions* (2) *has a single definite solution which exists in the interval I of continuity of the coefficients of the system*; *also, the solution can be obtained by the method of successive approximations.*

This result is also valid in the case when I is an open interval $c < x < d$, since, by what has been proved, we have the existence and uniqueness of the solution in every finite closed interval $a \leqslant x \leqslant b$, which contains the initial point x_0 and lies inside the interval $c < x < d$.

We might equally have considered a non-homogeneous system, i.e. with the extra functions $f_1(x)$, $f_2(x)$, continuous in the interval I, added to the right-hand sides of equations (1). The foregoing proof remains in force here.

The linear equation of the second order:

$$y'' + p(x)\,y' + q(x)\,y = 0 \tag{15}$$

can be written in the form of a system, on introducing an unknown $z = y'$ in addition to y:

$$\frac{dy}{dx} = z; \quad \frac{dz}{dx} = -\,p(x)\,z - q(x)\,y,$$

The result stated above is therefore also justified for (15), with the initial conditions

$$y|_{x=x_0} = y_0; \quad y'_{x=x_0} = y'_0, \tag{16}$$

in the interval I of continuity of the coefficients $p(x)$, $q(x)$.

We can use conditions (16) to re-write equation (15) as

$$y = y_0 + y'_0 x - \int\limits_{x_0}^{x} dx \int\limits_{x_0}^{x} [p(x)\,y' + q(x)\,y]\,dx, \tag{17}$$

where we can replace the double by a single integral in accordance with expression (23) of [15]. Equation (17) makes it possible to apply the method of successive approximations to (15) without reducing the equation to a system.

Example. We apply the method of successive approximations to the example of [46]:

$$y'' - xy = 0,$$

We take the initial conditions $y|_{x=0} = 1$ and $y'|_{x=0} = 0$. Equation (17) now becomes:

$$y = 1 + \int\limits_0^x dx \int\limits_0^x xy\,dx.$$

On substituting $y = 1$ on the right-hand side, we get the second approximation:

$$y_1(x) = 1 + \int\limits_0^x dx \int\limits_0^x x\,dx = 1 + \frac{x^3}{2\cdot 3}\,.$$

The third approximation is:

$$y_2(x) = 1 + \int\limits_0^x dx \int\limits_0^x x\left(1 + \frac{x^3}{2\cdot 3}\right) dx = 1 + \frac{x^3}{2\cdot 3} + \frac{x^6}{2\cdot 3\cdot 5\cdot 6}\,.$$

On passing to the limit, we clearly get the power series:

$$y = 1 + \frac{1}{3!}x^3 + \frac{1\cdot 4}{6!}x^6 + \frac{1\cdot 4\cdot 7}{9!}x^9 + \ldots,$$

which we had in [46].

51. The case of a non-linear equation. The existence and uniqueness theorem may also be proved for non-linear equations by using successive approximations, though the final result is stated somewhat differently. We shall take for simplicity a first order equation:

$$y' = f(x, y) \qquad (18)$$

with the initial condition:

$$y|_{x=x_0} = y_0. \qquad (19)$$

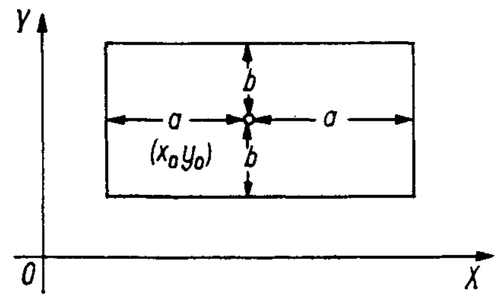

FIG. 27

We suppose that the given function $f(x, y)$ is continuous in the neighbourhood of the initial point (x_0, y_0) and has a bounded derivative with respect to y in this neighbourhood. More rigorously, a rectangle Q exists in the xy plane (Fig. 27):

$$\left.\begin{aligned} x_0 - a \leqslant x \leqslant x_0 + a; \\ y_0 - b \leqslant y \leqslant y_0 + b, \end{aligned}\right\} \qquad (20)$$

in which $f(x, y)$ is continuous, has a partial derivative with respect to y, and where

$$\left|\frac{\partial f(x,y)}{\partial y}\right| < k, \qquad (21)$$

k being a definite positive number. We can show, as in the case of a linear equation, that (18) with initial condition (19) is equivalent to the equation:

$$y = y_0 + \int_{x_0}^{x} f[t, y(t)]\,\mathrm{d}t. \qquad (22)$$

We assume here that the interval of variation of x does not exceed $(x_0 - a,\ x_0 + a)$, whilst the value of the continuous function $y(x)$ does not lie outside $(y_0 - b,\ y_0 + b)$, i.e. we assume that the point with abscissa x and ordinate $y(x)$ belongs to the rectangle Q.

Subtraction of successive approximations will lead to expressions similar to (4) and (5):

$$y_1(x) = y_0 + \int_{x_0}^{x} f(t, y_0)\,dt; \ldots;$$
$$y_n(x) = y_0 + \int_{x_0}^{x} f[t, y_{n-1}(t)]\,dt. \tag{23}$$

We turn to condition (21). If two points (x_1, y_1) and (x_1, y_2) of Q are taken with the same abscissae, we can use the formula of finite increments to write [I, 63]:

$$|f(x_1, y_2) - f(x_1, y_1)| = |y_2 - y_1| \left[\frac{\partial f(x_1, y)}{\partial y}\right]_{y=y_3},$$

where y_3 lies between y_1 and y_2. Condition (21) gives with this:

$$|f(x_1, y_2) - f(x_1, y_1)| < k|y_2 - y_1|. \tag{24}$$

This inequality is usually called a Lipschitz condition and is used for proving the convergence of $y_n(x)$ and the uniqueness of the solution. Let M be the greatest absolute value of the continuous function $f(x, y)$ in rectangle Q, i.e.

$$|f(x, y)| \leqslant M \quad [(x, y) \text{ in } Q]. \tag{25}$$

In carrying out the subtractions of (23), care must first of all be taken that the points with abscissae x and ordinates $y_n(x)$ do not pass outside rectangle Q, as defined by conditions (20). The first of these conditions gives the inequality $| x - x_0 | \leqslant a$ for x. The second leads to the inequality:

$$|y_n(x) - y_0| \leqslant b. \tag{26}$$

For this inequality to be satisfied for any n, x has to be subjected to the condition $| x - x_0 | \leqslant b/M$, as well as $| x - x_0 | \leqslant a$; thus we finally get for x:

$$|x - x_0| \leqslant a; \quad |x - x_0| \leqslant \frac{b}{M}. \tag{27}$$

We show that, with this, all the approximations satisfy inequality (26). The first of equations (23) gives

$$y_1(x) - y_0 = \int_{x_0}^{x} f(t, y_0)\,dt,$$

and if we evaluate the integral as usual, using (25), we get:

$$|y_1(x) - y_0| \leqslant M|x - x_0|,$$

hence, by the second of conditions (27), $|y_1(x) - y_0| \leqslant b$, so that (26) is satisfied for $n = 1$. Furthermore, the function $y_1(x)$ defined by the above expression is obviously continuous, given that conditions (27) are observed. Having noted these points, we can now pass on to find $y_2(x)$ with $n = 2$ in (23):

$$y_2(x) - y_0 = \int_{x_0}^{x} f[t, y_1(t)]\,dt,$$

whence, as above,

$$|y_2(x) - y_0| \leqslant M|x - x_0| \leqslant M\frac{b}{M} = b,$$

i.e. (26) is also satisfied for $n = 2$, and $y_2(x)$ is clearly continuous, given conditions (27), and so on. We can thus find successive approximations $y_n(x)$ in the interval $(x_0 - c,\ x_0 + c)$ where, by (27), c is the lesser of the two numbers a and b/M. Let us call this interval I. All the $y_n(x)$ are continuous in I, and in all future arguments we shall assume that x belongs to I.

We now consider $y_n(x) - y_{n-1}(x)$, and take $x - x_0 > 0$ for simplicity, as above. By (25), the first of equations (23) gives:

$$|y_1(x) - y_0| \leqslant M(x - x_0). \tag{28}$$

We take the second of equations (23) with $n = 2$ and subtract from the first:

$$y_2(x) - y_1(x) = \int_{x_0}^{x} \{f[t, y_1(t)] - f(t, y_0)\}\,dt,$$

whence [I, 95]

$$|y_2(x) - y_1(x)| \leqslant \int_{x_0}^{x} |f[t, y_1(t)] - f(t, y_0)|\,dt,$$

or, by (24):

$$|y_2(x) - y_1(x)| \leqslant \int_{x_0}^{x} k|y_1(t) - y_0|\,dt.$$

We obtain further, on using inequality (28):

$$|y_2(x) - y_1(x)| \leqslant kM \int_{x_0}^{x} (t - x_0)\,dt = kM\left[\frac{(t - x_0)^2}{2!}\right]_{t=x}^{t=x},$$

and finally,

$$| y_2(x) - y_1(x) | \leqslant kM \frac{(x - x_0)^2}{2!}. \tag{29}$$

We now write the second of equations (23) for $n = 2$ and $n = 3$ and find, on subtraction:

$$y_3(x) - y_2(x) = \int_{x_0}^{x} \{f[t, y_2(t)] - [t, y_1(t)]\} \mathrm{d}t.$$

This gives us, as above, on using (24) and (29):

$$| y_3(x) - y_2(x) | \leqslant k^2 M \frac{(x - x_0)^3}{3!}$$

and by proceeding in this way, we arrive at the general inequality:

$$| y_n(x) - y_{n-1}(x) | \leqslant \frac{M}{k} \frac{[k(x - x_0)]^n}{n!}. \tag{30}$$

If we replace $(x - x_0)$ on the right by its absolute value, the inequality is valid for all x of I. It follows from this inequality, as above, that in the interval I $y_n(x)$ tends uniformly with respect to x to a limit function $y(x)$, which is continuous and satisfies inequality (26), i.e. $| y(x) - y_0 | \leqslant b$. Hence, points with abscissae x and ordinates $y(x)$ belong to rectangle Q. We have by the continuity of $f(x, y)$:

$$\lim_{n \to \infty} f[t, y_{n-1}(t)] = f[t, y(t)] \qquad (t \text{ from } I).$$

The uniformity of this passage to the limit with respect to t in I is easily seen as follows. For any given positive ε, there exists a δ, by the uniform continuity of $f(x, y)$ in Q, such that $| f(x'', y'') - - f(x', y') | < \varepsilon$ if (x', y') and (x'', y'') are points of Q for which $| x'' - x' | < \delta$ and $| y'' - y' | < \delta$. Furthermore, since $y_{n-1}(t)$ tends uniformly to $y(t)$, there exists an N, the same for all t of I, such that $| y(t) - y_{n-1}(t) | < \delta$ for $n > N$ and all t of I. Hence it follows that for all t of I:

$$| f[t, y(t)] - f[t, y_{n-1}(t)] | < \varepsilon \qquad\qquad n > N,$$

which proves that the passage to the limit is uniform. We return to the second of equations (23) and pass to the limit on both sides with indefinite increase of n. We can pass to the limit under the integral sign, by the uniform convergence of $f[t, y_{n-1}(t)]$ to $f[t, y(t)]$, and we thus get equation (22) for the limit function.

The uniqueness remains to be proved. Let equation (22) have two solutions $y(x)$ and $Y(x)$ in some interval $(x_0 - d, x_0 + d)$ which lies

inside $(x_0 - a,\ x_0 + a)$, where d can be taken small enough for $y(x)$ and $Y(x)$ not to lie outside $(y_0 - b,\ y_0 + b)$. We substitute first one, then the other solution in (22) and subtract:

$$y(x) - Y(x) = \int_{x_0}^{x} \{f[t, y(t)] - f[t,\ Y(t)]\}\,dt,$$

whence

$$|y(x) - Y(x)| \leqslant \int_{x_0}^{x} |f[t, y(t)] - f[t,\ Y(t)]|\,dt \qquad (t \geqslant x_0),$$

and by (24):

$$|y(x) - Y(x)| \leqslant k \int_{x_0}^{x} |y(t) - Y(t)|\,dt.$$

We take an interval of length l_1 such that $kl_1 = \theta$ is less than unity and show, as previously, that $y(x)$ and $Y(x)$ coincide. Hence, *with the assumptions made regarding* $f(x, y)$, *equation* (18) *with initial conditions* (19) *has a definite solution which exists in the interval* $(x_0 - c,\ x_0 + c)$, *where c is the lesser of the numbers a and* b/M, *and which can be obtained by the method of successive approximations.* We notice that the definition of the interval in which x varies is more complicated for a non-linear equation than for a system of linear equations, where it is simply the interval of continuity of the coefficients. We explain the matter more precisely by an example.

Example. We take the equation:

$$y' = x + y^2 \tag{31}$$

with the initial condition

$$y|_{x=0} = 0. \tag{32}$$

Equation (22) becomes:

$$y(x) = \int_0^x [t + y^2(t)]\,dt. \tag{33}$$

We replace $y(t)$ on the right by zero and find the second approximation:

$$y_1(x) = \int_0^x t\,dt = \frac{x^2}{2}.$$

We get the third approximation by substituting this in the right-hand side of (33):

$$y_2(x) = \int_0^x \left[t + \frac{t^4}{4}\right] dt = \frac{x^2}{2} + \frac{x^5}{20}$$

and so on.

We now consider how to define the interval of variation of x in which we may apply the method of successive approximations. The right-hand side of equation (31) is continuous and has a bounded derivative with respect to y in any rectangle drawn about the point (0, 0), so that the a and b in conditions (20) can be taken arbitrarily. With this, $M = \max |x + y^2| = a + b^2$, and inequalities (26), defining the required interval of variation of x, become

$$|x| \leqslant a; \quad |x| \leqslant \frac{b}{a + b^2}.$$

If b is taken as either large or near zero, the second inequality results in a very narrow interval. The same is true if a is taken as large, whereas small a also gives a narrow interval, by the first inequality. It follows that we cannot have as large an interval as we please for x, although the right-hand side of (31) has no singularities for finite x and y.

52. Singular points of first order differential equations. If the right-hand side of the equation

$$y' = f(x, y) \tag{34}$$

is continuous and has a bounded derivative with respect to y at and in the neighbourhood of the point (x_0, y_0), the existence and uniqueness theorem shows that one, and only one, integral curve passes through this point. A point at which the above conditions are not satisfied by $f(x, y)$ is called a *singular point of the equation*, and the existence and uniqueness theorem is no longer valid at such a point.

We re-write (34) in a form containing the differentials:

$$\frac{\mathrm{d}x}{P(x, y)} = \frac{\mathrm{d}y}{Q(x, y)}, \tag{35}$$

and for simplicity, we take $P(x, y)$ and $Q(x, y)$ as polynomials in x and y. If $P(x_0, y_0) \neq 0$, (35) can be written as

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{Q(x, y)}{P(x, y)},$$

and given this condition only, the right-hand side is continuous at and in the neighbourhood of the point (x_0, y_0) and has a bounded derivative with respect to y, as found by the usual rule for differentiation of a fraction. The conditions for the existence and uniqueness theorem are thus satisfied at (x_0, y_0) if $P(x_0, y_0) \neq 0$, and one, and only one, integral curve of equation (35) passes through this point. If $P(x_0, y_0) = 0$ but $Q(x_0, y_0) \neq 0$, (35) can be written as

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{P(x, y)}{Q(x, y)},$$

with x taken as a function of y. The denominator of the right-hand side does not vanish at (x_0, y_0), and we can see, as above, that the existence and uniqueness theorem is applicable at this point. The singular points of equation (35) are thus the points at which $P(x, y)$ and $Q(x, y)$ vanish simultaneously, their coordinates being found as the real solutions of the system:

$$P(x,y)=0; \quad Q(x,y)=0. \tag{36}$$

What has been said also applies to the case when $P(x, y)$ and $Q(x, y)$ are series, expanded in positive integral powers of $(x-x_0)$ and $(y-y_0)$. If the constant term of at least one series differs from zero, the existence and uniqueness theorem is applicable at (x_0, y_0); otherwise, we have a singular point of the equation.

The idea of a singular point may be explained by the example of steady fluid flow [12]. Let $P(x, y)$, $Q(x, y)$ be the projections of the velocity vector $\mathbf{v}(x, y)$ on the coordinate axes. Equation (35) expresses the condition that the tangent is parallel to the velocity vector and, in fact, is the differential equation of a stream line. At a point where $\mathbf{v}(x, y)$ differs from zero, at least one of its projections $P(x, y)$ and $Q(x, y)$ must differ from zero, and one, and only one, stream line passes through the point, by the existence and uniqueness theorem. On the other hand, points where $\mathbf{v}(x, y)$ vanishes, i.e. where equations (36) apply, are singular points of (35) and are called critical points of the flow in question. A different situation is obtained here: stream lines can intersect at the point, approach the point asymptotically or form a closed curve round it. Singular points can thus be of various kinds, and it is important to know which kind when studying the motion (the integral curves of the equation). We deal with the problem in a particular example in the next section.

53. The stream lines of collinear plane fluid motion. We take the particular case when the projections of the velocity are first degree polynomials:

$$P(x,y)=a_{11}x+a_{12}y+b_1; \quad Q(x,y)=a_{21}x+a_{22}y+b_2;$$

the fluid flow being spoken of as collinear in this case.

We start by assuming that the straight lines

$$a_{11}x+a_{12}y+b_1=0 \quad \text{and} \quad a_{21}x+a_{22}y+b_2=0 \tag{37}$$

are not parallel. The constant terms b_1 and b_2 vanish, on taking the point of intersection of the lines as the origin. The equation of motion

becomes

$$\frac{dx}{a_{11}x + a_{12}y} = \frac{dy}{a_{21}x + a_{22}y}, \tag{38}$$

for which the origin, $x = y = 0$, is obviously a singular point. We show how the nature of the singular point is revealed by the form of the coefficients a_{ik}.

Clearly, (38) is a homogeneous equation and can be integrated by the method given in [3]. We use another method, however, which consists in introducing new variables ξ and η, so as first to reduce (38) to a more convenient form for direct inspection.

We put

$$\xi = m_1 x + n_1 y; \quad \eta = m_2 x + n_2 y, \tag{39}$$

whence

$$d\xi = m_1\, dx + n_1\, dy; \quad d\eta = m_2\, dx + n_2\, dy.$$

We get from (38), on eliminating dx and dy:

$$\frac{d\xi}{m_1(a_{11}x + a_{12}y) + n_1(a_{21}x + a_{22}y)} = \frac{d\eta}{m_2(a_{11}x + a_{12}y) + n_2(a_{21}x + a_{22}y)}. \tag{40}$$

We now define the coefficients in (39) in such a way that the denominators of the fractions written are proportional to ξ and η respectively. The first denominator gives us

$$m_1(a_{11}x + a_{12}y) + n_1(a_{21}x + a_{22}y) = \varrho(m_1 x + n_1 y),$$

whence, on comparing coefficients of x and y, we obtain a system of homogeneous equations for m_1 and n_1:

$$\left.\begin{aligned} (a_{11} - \varrho)\, m_1 + a_{21}n_1 &= 0 \\ a_{12}m_1 + (a_{22} - \varrho)\, n_1 &= 0. \end{aligned}\right\} \tag{41_1}$$

Similarly, on equating the second denominator to $\varrho\eta$, we get the system for m_2 and n_2:

$$\left.\begin{aligned} (a_{11} - \varrho)\, m_2 + a_{21}n_2 &= 0 \\ a_{12}m_2 + (a_{22} - \varrho)\, n_2 &= 0, \end{aligned}\right\} \tag{41_2}$$

where the coefficient of proportionality ϱ will now have a different value.

We cannot have $m = n = 0$, since the change of variables of (39) then becomes meaningless. Systems (41_1) and (41_2) must thus have

solutions differing from zero. But two homogeneous linear equations

$$\alpha_1 x + \beta_1 y = 0; \quad \alpha_2 x + \beta_2 y = 0$$

have a solution differing from $x = y = 0$ when, and only when, the straight lines corresponding to the equations coincide, i.e. when their coefficients are proportional. With systems (41_1) and (41_2), this amounts to the condition

$$\frac{a_{11} - \varrho}{a_{12}} = \frac{a_{21}}{a_{22} - \varrho},$$

leading to the quadratic equation for ϱ:

$$\varrho^2 - (a_{11} + a_{22})\,\varrho + (a_{11}a_{22} - a_{12}a_{21}) = 0. \tag{42}$$

With this, systems (41_1) and (41_2) both reduce to a single equation, and we can determine the solution differing from $m = n = 0$.

We now consider in detail the various possible cases.

(A) *Equation* (42) *has two distinct roots* ϱ_1 *and* ϱ_2.

On substituting $\varrho = \varrho_1$ in (41_1) and $\varrho = \varrho_2$ in (41_2), we can determine the coefficients in (39), as described above; equation (40) then reduces to an equation with separable variables:

$$\frac{d\xi}{\varrho_1 \xi} = \frac{d\eta}{\varrho_2 \eta}. \tag{43}$$

It may be shown that, with this, equations (39) are soluble with respect to x and y.

We proceed by considering various particular cases of case A.

(1) *The roots* ϱ_1 *and* ϱ_2 *of* (42) *are real and have the same sign.* Integration of (43) gives us:

$$\log \xi^{\varrho_2} = \log \eta^{\varrho_1} + \log C_1,$$

where $\log C_1$ is taken as the arbitrary constant.

Hence:

$$\xi^{\varrho_2} = C_1 \eta^{\varrho_1}; \quad \xi = C\eta^{\frac{\varrho_1}{\varrho_2}} \qquad \left(C = C_1^{\frac{1}{\varrho_2}}\right),$$

or

$$(m_1 x + n_1 y) = C\,(m_2 x + n_2 y)^{\frac{\varrho_1}{\varrho_2}}. \tag{44}$$

The fraction ϱ_1/ϱ_2 is positive in the present case, so that the coordinates $x = y = 0$ satisfy (44) for any value of C, i.e. every stream

line (integral curve) has a singular point (Fig. 28). A point of this sort is called a *node*.

(2) *Roots ϱ_1 and ϱ_2 real and of different sign.* Here, ϱ_1/ϱ_2 is negative, and we write it as $(-\mu)$, where μ is a positive number; the general solution (44) now becomes:

$$(m_1 n + x_1 y)(m_2 x + n_2 y)^\mu = C \qquad (\mu > 0). \tag{45}$$

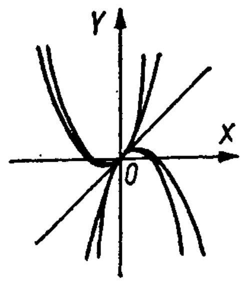

FIG. 28

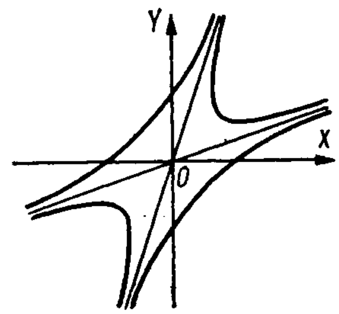

FIG. 29

We obtain $C = 0$ on substituting $x = y = 0$, so that the equation of stream lines passing through the origin runs:

$$(m_1 x + n_1 y)(m_2 x + n_2 y)^\mu = 0$$

corresponding to which we have the two straight lines:

$$\begin{aligned} m_1 x + n_1 y &= 0; \\ m_2 x + n_2 y &= 0. \end{aligned} \tag{46}$$

It follows that in this case two and only two stream lines (integral curves) pass through the singular point, which is here referred to as a *neutral* or *saddle point*. The curves (45) are similar to hyperbolas for values of C differing from zero (or actually are hyperbolas for $\mu = 1$), whilst the straight lines (46) are their asymptotes (Fig. 29).

(3) *Roots ϱ_1 and ϱ_2 complex conjugates, with non-zero real parts:*

$$\varrho_1 = \alpha + \beta i; \quad \varrho_2 = \alpha - \beta i \quad (\alpha \text{ and } \beta \neq 0).$$

If we substitute conjugate values of ϱ in the coefficients of systems (41_1) and (41_2), we get systems whose corresponding coefficients are conjugates. Replacing i by $(-i)$ in any solution m_1, n_1 of one system therefore gives us the solution m_2, n_2 of the other system. It follows from (39) that ξ and η can also be taken as conjugates:

$$\xi = \xi_1 + \eta_1 i; \quad \eta = \xi_1 - \eta_1 i,$$

where ξ_1 and η_1 are real polynomials in x and y of the form

$$\xi_1 = p_1 x + q_1 y; \quad \eta_1 = p_2 x + q_2 y. \tag{47}$$

Equation (43) becomes:

$$\frac{\mathrm{d}\xi_1 + i\,\mathrm{d}\eta_1}{(\alpha + \beta i)(\xi_1 + \eta_1 i)} = \frac{\mathrm{d}\xi_1 - i\,\mathrm{d}\eta_1}{(\alpha - \beta i)(\xi_1 - \eta_1 i)}$$

or

$$\frac{\mathrm{d}\xi_1 + i\mathrm{d}\eta_1}{(\alpha\xi_1 - \beta\eta_1) + (\beta\xi_1 + \alpha\eta_1)\,i} = \frac{\mathrm{d}\xi_1 - i\,\mathrm{d}\eta_1}{(\alpha\xi_1 - \beta\eta_1) - (\beta\xi_1 + \alpha\eta_1)\,i}\,.$$

We obtain in the usual way, on forming new proportions, by first adding numerators and denominators, then subtracting them:

$$\frac{\mathrm{d}\xi_1}{\alpha\xi_1 - \beta\eta_1} = \frac{\mathrm{d}\eta_1}{\beta\xi_1 + \alpha\eta_1}\,,$$

whence

$$\xi_1\,\mathrm{d}\xi_1 + \eta_1\,\mathrm{d}\eta_1 = \frac{\alpha}{\beta}\,(\xi_1\,\mathrm{d}\eta_1 - \eta_1\,\mathrm{d}\xi_1)$$

or

$$\frac{\xi_1\,\mathrm{d}\xi_1 + \eta_1\,\mathrm{d}\eta_1}{\xi_1^2 + \eta_1^2} = \frac{\alpha}{\beta}\,\frac{1}{1 + \frac{\eta_1^2}{\xi_1^2}}\cdot\frac{\eta_1\,\mathrm{d}\eta_1 - \eta_1\,\mathrm{d}\xi_1}{\xi_1^2}\,.$$

On writing

$$u = \xi_1^2 + \eta_1^2; \quad v = \frac{\eta_1}{\xi_1}\,,$$

we get

$$\frac{\mathrm{d}u}{2u} = \frac{\alpha}{\beta}\,\frac{\mathrm{d}v}{1 + v^2}\,; \quad \frac{1}{2}\log u = \frac{\alpha}{\beta}\arctan v + \log C$$

and the general solution is therefore:

$$\log\sqrt{\xi_1^2 + \eta_1^2} = \frac{\alpha}{\beta}\arctan\frac{\eta_1}{\xi_1} + \log C \quad \text{or} \quad \sqrt{\xi_1^2 + \eta_1^2} = C\mathrm{e}^{\frac{\alpha}{\beta}\arctan\frac{\eta_1}{\xi_1}} \tag{48}$$

We obtain

$$r = C\mathrm{e}^{\frac{\alpha}{\beta}\theta}\,,$$

on taking polar coordinates $\xi_1 = r\cos\theta$ and $\eta_1 = r\sin\theta$ in the (ξ_1, η_1) plane, i.e. the stream lines are logarithmic spirals in (ξ_1, η_1) coordinates, approaching the origin from the same direction [I, 83].

The spiral forms will be maintained on passing to the original (x, y) coordinates, related to the (ξ_1, η_1) coordinates by (47). It follows that no stream line passes through the singular point in the present case, whereas they all approach it asymptotically in spirals (Fig. 30). A singular point of this kind is called a *focus*.

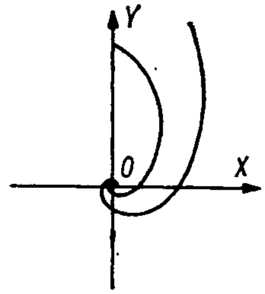

FIG. 30

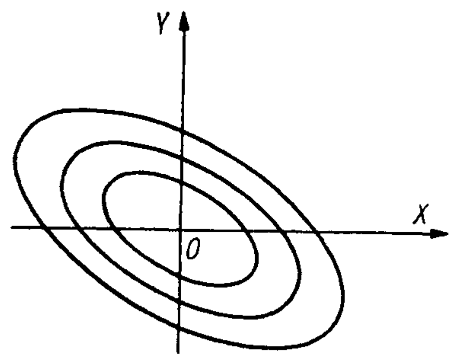

FIG. 31

(4) *Roots* ϱ_1 *and* ϱ_2 *pure imaginary* $(\pm\beta\, i)$. We obtain on putting $a = 0$ in (48):

$$\xi_1^2 + \eta_1^2 = C^2, \tag{49}$$

or in the original coordinates:

$$(p_1 x + q_1 y)^2 + (p_2 x + q_2 y)^2 = C^2. \tag{50}$$

We get similar ellipses instead of circles (49). None of the stream lines (integral curves) pass through the singular point in this case; they form closed loops about the point (Fig. 31), instead of spiralling round it as in the previous case. A singular point of this sort is called a *centre*.

(*B*) *Equation* (42) *has a double, non-zero root* $\varrho_1 = \varrho_2$. Two cases are possible on substituting $\varrho = \varrho_1$ in the coefficients of systems (41_1) or (41_2): either all the coefficients vanish, or there is at least one coefficient differing from zero. We have in the first case:

$$a_{12} = a_{21} = 0; \quad a_{11} = a_{22} = \varrho_1, \tag{51}$$

and system (38) takes the form:

$$\frac{dx}{\varrho_1 x} = \frac{dy}{\varrho_1 y} \quad \text{or} \quad \frac{dx}{x} = \frac{dy}{y};$$

its general solution is a family of straight lines passing through the origin, i.e. the origin is a node.

At least one of the coefficients

$$a_{12}, a_{21}, a_{11} - \varrho_1, a_{22} - \varrho_1$$

differs from zero. It is easily shown that a_{12} and a_{21} cannot both be zero here; for, if $a_{12} = a_{21} = 0$, we get $a_{11} = a_{22} = \varrho_1$ by the fact that ϱ_1 is a double root of (42); equation (42) must reduce to $\varrho^2 - - (a_{11} + a_{22})\, \varrho + a_{11} a_{22} = 0$ on the assumption made, and the condition for a double root gives $a_{11} = a_{22}$ from this equation, the common value of a_{11} and a_{22} being the value of the double root. Conditions (51) are therefore satisfied if we take $a_{12} = a_{21} = 0$, and this contradicts our hypothesis. Thus at least one of the coefficients a_{12} or a_{21} differs from zero. Suppose, say, $a_{21} \neq 0$. The double root of (42) is now clearly:

$$\varrho_1 = \frac{a_{11} + a_{22}}{2},$$

and on substituting $\varrho = \varrho_1$ system (41_1) must reduce to a single equation, as remarked above:

$$\frac{a_{11} - a_{22}}{2} m_1 + a_{21} n_1 = 0.$$

We take $m_1 = a_{21}$ and $n_1 = -(a_{11} - a_{22})/2$, i.e.

$$\xi = a_{21} x - \frac{a_{11} - a_{22}}{2} y, \tag{52}$$

and retain y as the second variable. The differential equations can be written as

$$\frac{\mathrm{d}\xi}{\varrho_1 \xi} = \frac{\mathrm{d}y}{a_{21} x + a_{22} y}$$

or, on replacing x by its expression in accordance with (52):

$$\frac{\mathrm{d}\xi}{\varrho_1 \xi} = \frac{\mathrm{d}y}{\xi + \varrho_1 y}.$$

On introducing a new variable t instead of y:

$$y = t\xi,$$

we arrive at the equation

$$\frac{1}{\varrho_1} \frac{\mathrm{d}\xi}{\xi} = \mathrm{d}t,$$

and integration gives us the general solution:

$$y = \frac{\xi}{\varrho_1} \log (C\xi).$$

The sign of ξ must be the same as the sign of C, whilst obviously $y \to 0$ as $\xi \to 0$; in addition,

$$y' = \frac{1}{\varrho_1}[1 + \log(C\xi)] \to \infty,$$

i.e. the integral curves in (ξ, y) coordinates meet at the origin, where they touch the OY axis (Fig. 32), so that the origin is a node.

A basic assumption in the transformation of equation (35) to the form (38) was that the straight lines (37) are not parallel. If these

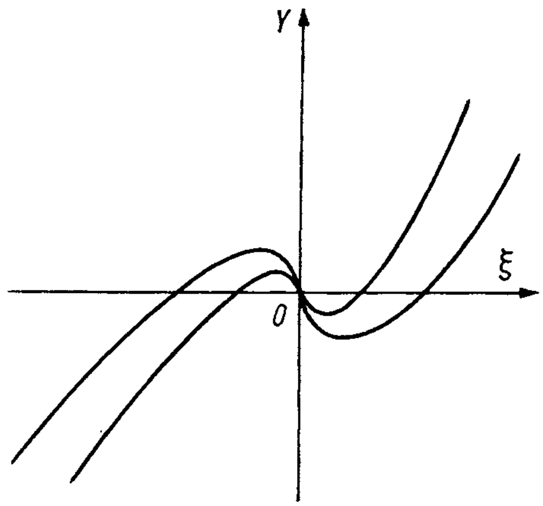

FIG. 32

are parallel, their left-hand sides do not vanish identically, and the differential equation of the stream lines:

$$\frac{\mathrm{d}x}{a_{11} + a_{12}\,y + b_1} = \frac{\mathrm{d}y}{a_{21}\,x + a_{22}\,y + b_2}$$

has no singular points; now, one and only one stream line passes through every point of the plane.

In a more general case than that of collinear motion, the equation with a singular point at the origin has the form:

$$\frac{\mathrm{d}x}{a_{11}\,x + a_{12}\,y + b_1\,x^2 + c_1\,xy + \dots} = \frac{\mathrm{d}y}{a_{21}\,x + a_{22}\,y + b_2\,x^2 + c_2\,xy + \dots}\,. \tag{53}$$

The denominators here contain terms of higher order than the first in x and y. Integration of such equations does not lead to a quadrature except in special cases, though the nature of the singular point can sometimes be found from the initial coefficients in the denominators of (53) without in fact carrying out any integrations.

The argument runs roughly as follows: for x and y of small absolute value, i.e. in the neighbourhood of the origin, the higher order terms in the denominators of (53) will be small compared with the first order terms and the form of the integral curves near the origin may be supposed to be the same as it would be if we only considered the first order terms in the denominators. If this is so, we encounter the same sorts of singular point with equations (53) as were obtained above for collinear motion. These suppositions are correct, with certain exceptions. We state the final result, without proof:

(1) If the roots of quadratic equation (42) are real and distinct and have the same sign, the singular point $x = y = 0$ of equation (53) is a node. This means that every integral curve that approaches sufficiently close to the singular point strikes this point.

(2) If the roots of (42) are real and have opposite signs, the singular point is a saddle point, i.e. two integral curves pass through it.

(3) If the roots of (42) are complex, with non-zero real parts, the singular point is a focus.

(4) If the roots of (42) are pure imaginary, the singular point is a focus or centre.

CHAPTER III

MULTIPLE AND LINE INTEGRALS. IMPROPER INTEGRALS. INTEGRALS THAT DEPEND ON A PARAMETER

§ 6. Multiple integrals

54. Volumes. We have so far considered the definite integral

$$\int_a^b f(x)\,\mathrm{d}x,$$

as the limit of a sum, when the function $f(x)$ is defined over the segment (a, b) of axis OX. In other words, the domain of integration has always been a straight line.

We generalize the concept of integral in the present article to include the cases when integration is over a plane domain, or a domain in space, or finally, over a surface. We shall make use of intuitive ideas of area and volume for the present, and shall not dwell on certain points arising in connection with passage to the limit. The reader can find the main points of a rigorous treatment in a later section of this chapter. We start with double integrals, which have the same sort of connection with the calculation of volume as the integral written above has with the calculation of area. We shall consider the calculation of volume as a preliminary to our treatment of double integrals.

We know that the area bounded by the curve $y = f(x)$, the x axis, and the ordinates $x = a$, $x = b$, can be evaluated with the aid of a definite integral, and is in fact expressed by the integral written above [I, 87].

We turn to the analogous problem regarding the volume v of a body bounded by a given surface (S), whose equation is

$$z = f(x, y), \tag{1}$$

by the XOY plane, and by the cylinder (C) with generators parallel to the z axis which project (S) into the area (σ) in the XOY plane (Fig. 33).

We reduced the calculation of the volume of a body to a definite integral in [I, 104], where all we required to know were parallel sections through the body. We use this method in the present problem.

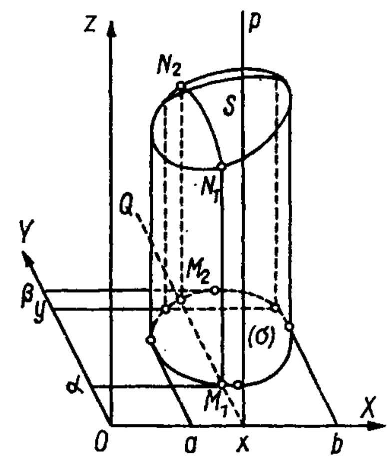

FIG. 33

We suppose for simplicity that the surface (S) lies wholly above the XOY plane and that the contour (l) bounding (σ) is intersected at two points only by lines drawn parallel to the coordinate axes.

We dissect the body by means of planes parallel to the YOZ plane, which cut the XOY plane in straight lines parallel to the y axis (Figs. 33 and 34). Let the abscissae of the extreme cutting planes be a, b. These are also the abscissae of the points that divide the contour into two parts (1) and (2), these

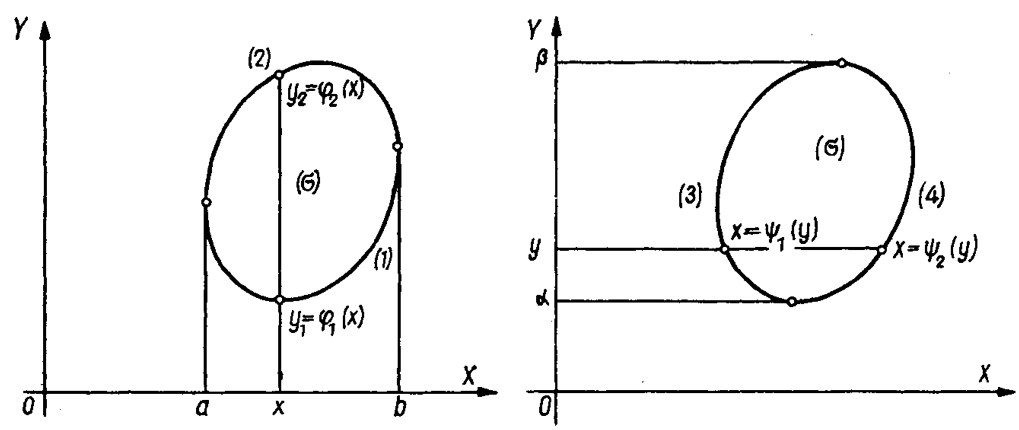

FIG. 34

parts being the loci of the points of entry and exit respectively of lines parallel to the y axis into the domain (σ) (Fig. 34). Each part has its own equation:

$$y_1 = \varphi_1(x); \quad y_2 = \varphi_2(x). \tag{2}$$

The area of the section of the body cut by the plane PQ, distant x from YOZ, depends on x. Let this area be $S(x)$; we have [I, 104]:

$$v = \int_a^b S(x)\,dx. \tag{3}$$

It remains to find an expression for function $S(x)$, which is the area of the figure $M_1 N_1 N_2 M_2$, lying in the plane PQ, and bounded by the curve of intersection N_1N_2 of plane PQ and the surface (S), by $M_1 M_2$, parallel to OY, and by the two ordinates M_1N_1 and $M_2 N_2$.

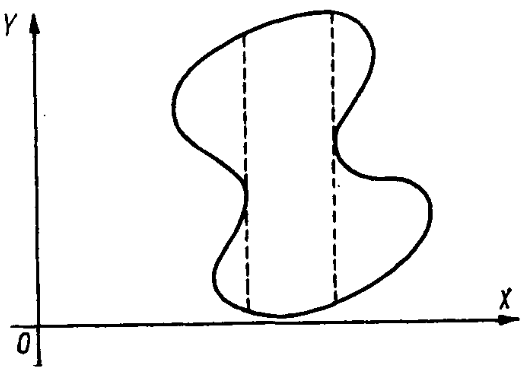

FIG. 35

Since x is constant for all points of the section concerned, the ordinate of the curve N_1N_2 can be taken as a function of y, given by the equation:

$$z = f(x, y)$$

with constant x; with this, the independent variable y varies in the interval (y_1, y_2), where y_1 and y_2 are the ordinates of the points of entry and departure respectively of the straight line $M_1 M_2$ into or from the domain (σ).

We can write by [I, 87]:

$$S(x) = \int_{y_1}^{y_2} f(x, y)\,dy;$$

and on substituting in (3), we have:

$$v = \int_a^b dx \int_{y_1}^{y_2} f(x, y)\,dy. \tag{4}$$

Here we have expressed the volume as an *iterated integral*, where integration is first carried out with respect to y with constant x, then the result obtained integrated with respect to x.

If we dissect the given body with planes parallel to XOZ, we can express the same volume as

$$v = \int_\alpha^\beta dy \int_{x_1}^{x_2} f(x, y)\,dx, \tag{5}$$

where x_1 and x_2 are known functions of y:

$$x_1 = \psi_1(y); \quad x_2 = \psi_2(y), \tag{6}$$

and α and β denote the extreme values of y on the contour (l) (Figs. 33 and 34).

Two assumptions were made in deducing expressions (4) and (5): (1) that the surface (S) lies wholly above the XOY plane, and (2) that the contour (l) of the projection (σ) of (S) on the XOY plane is intersected at two points only by any line parallel to a coordinate axis. If condition (1) is not satisfied, the right-hand sides of (4) and (5), instead of giving the volumes, give the *algebraic sum of the volumes*, volumes being considered (+) or (−) depending on whether they lie above or below the XOY plane. If condition (2) is not satisfied, as for instance in Fig. 35, where contour (l) is intersected at more than one pair of points by a line $x = \text{const.}$, the domain (σ) must be divided into sub-domains, in each of which condition (2) is satisfied. The surface (S) and volume v are now correspondingly subdivided, and each partial volume can be calculated by means of (4).

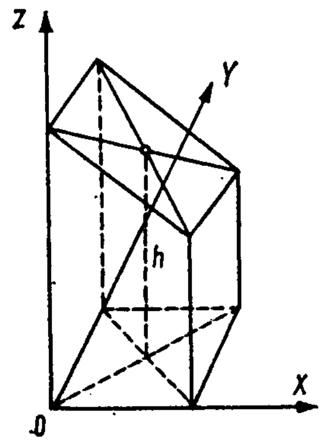

FIG. 36

Examples. **1.** The volume of a truncated rectangular prism (Fig. 36). The base is formed by the x and y axes and by the straight lines $x = k$, $y = l$. The cutting plane has the equation:

$$\frac{x}{\lambda} + \frac{y}{\mu} + \frac{z}{\nu} = l.$$

Expression (4) becomes in this case:

$$v = \int_0^k dx \int_0^l z\,dy = \int_0^k dx \int_0^l \nu\left(1 - \frac{x}{\lambda} - \frac{y}{\mu}\right) dy = \nu \int_0^k dx \left(y - \frac{xy}{\lambda} - \frac{y^2}{2\mu}\right)\Bigg|_{y=0}^{y=l} =$$

$$= \nu \int_0^k \left(l - \frac{xl}{\lambda} - \frac{l^2}{2\mu}\right) dx = \nu\left(kl - \frac{k^2 l}{2\lambda} - \frac{kl^2}{2\mu}\right) = kl\,\nu\left(1 - \frac{k}{2\lambda} - \frac{l}{2\mu}\right) = \sigma h,$$

where σ is the base area and h is the ordinate of the point of intersection of the diagonals of the upper section (corresponding to $x = k/2$, $y = l/2$).

2. The volume of the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

If the ellipsoid is cut by planes $z = \text{const.}$, ellipses are obtained with semi-axes

$$a\sqrt{1 - \frac{z^2}{c^2}}, \quad b\sqrt{1 - \frac{z^2}{c^2}}$$

and with area

$$S(z) = \pi ab\left(1 - \frac{z^2}{c^2}\right),$$

so that the required volume is

$$v = \int_{-c}^{+c} \pi ab\left(1 - \frac{z^2}{c^2}\right) dz = \frac{4}{3}\pi abc.$$

55. Double integrals. We obtained an approximate expression for the area of the curve $y = f(x)$ [I, 87] by dividing it into vertical strips and taking the area of each as equal to that of a rectangle with the same base and with height equal to some mean value of the ordinate of the curve in the strip. As the number of strips increased and the width of each tended to zero, the error tended to zero, and the approximate formula became the definite integral in the limit, accurately expressing the area.

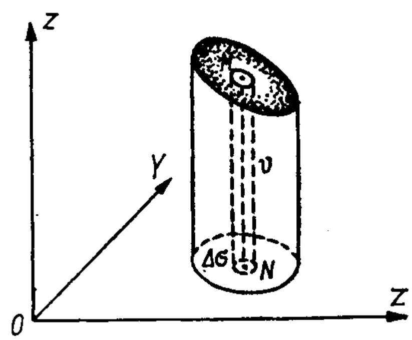

FIG. 37

An analogous method can be used for calculating volumes. We divide the domain (σ) (Fig. 37) into a large number of small elements $\Delta\sigma$ of arbitrary shape. Here, $\Delta\sigma$ refers either to the element itself or to its area. We take each element as the base of a cylinder which cuts out an elementary volume from the volume v on being continued to its intersection with surface (S). The elementary volume can evidently be taken as approximately equal to the volume of the cylinder of base $\Delta\sigma$ and of height equal to the ordinate, i.e. the value of z, of any point of the element of surface whose projection is $\Delta\sigma$. In other words, if we take any point N of the element $\Delta\sigma$ and let the ordinate of the corresponding point M of the surface (S), i.e. the value of $f(x, y)$ at the point N, be denoted for brevity by $f(N)$, we have $f(N)\Delta\sigma$ for the elementary volume, and

$$v \sim \sum_{(\sigma)} f(N)\,\Delta\sigma,$$

where summation extends over all the elements $\Delta\sigma$ that make up the area (σ).

The smaller the elements $\Delta\sigma$ and the larger their number, the closer the approximation obtained, and we can write in the limit:

$$\lim \sum_{(\sigma)} f(N)\,\Delta\sigma = v.$$

Turning aside from geometric concepts, we can define the above limit of a sum independently of the geometric form of the function $f(N)$; the limit is called *the double integral of function $f(N)$ over the domain (σ), and is written as:*

$$\iint_{(\sigma)} f(N)\,\mathrm{d}\sigma = \lim \sum_{(\sigma)} f(N)\,\Delta\sigma.$$

The existence of the limit written is self-evident, since, as we have noted, it necessarily gives the volume v described above. This is not a rigorous argument, of course, but a rigorous analytic proof of the existence of the limit is possible for fairly general conditions regarding $f(N)$ and for continuous functions without exception.

On setting $f(N) = 1$, we can express the area σ of the domain (σ) as a double integral:

$$\sigma = \iint_{(\sigma)} \mathrm{d}\sigma\,.$$

We give the detailed definition of a double integral: let (σ) be a bounded plane domain, and let $f(N)$ be a function of points of this domain, i.e. a function that takes a definite value at every point N of (σ). We divide (σ) into n sub-domains with areas $\Delta\sigma_1, \Delta\sigma_2, \ldots, \Delta\sigma_n$, and let $N_1, N_2, \ldots, N_n$ be arbitrary points of the sub-domains. We form the sum of the products:

$$\sum_{k=1}^{n} f(N_k)\cdot\Delta\sigma_k.$$

On indefinite increase in the number n of sub-domains and indefinite decrease of the area $\Delta\sigma_k$ of each, the limit of the sum is called the double integral of function $f(N)$ over domain (σ):

$$\iint_{(\sigma)} f(N)\,\mathrm{d}\sigma = \lim \sum_{k=1}^{n} f(N_k)\,\Delta\sigma_k.$$

Remark. Let d_k be the maximum distance between two points of the sub-domain of area $\Delta\sigma_k$ (the diameter of the domain), and let d be the greatest of the numbers $d_1, d_2, \ldots, d_n$. The indefinite decrease of each of the $\Delta\sigma_k$ referred to in the definition means that $d \to 0$.

If the magnitude of the integral is denoted by I, the above definition is equivalent to the following: given any positive ε, there exists a positive η such that [cf. I, 87]

$$\left| I - \sum_{k=1}^{n} f(N_k)\,\Delta\sigma_k \right| \leqslant \varepsilon,$$

provided $d \leqslant \eta$. A full account of the theory of multiple integrals appears at the end of the chapter and includes a rigorous definition of area, a more precise description of the domain (σ) over which integration can be carried out, as well as an explanation of how dissection is possible into sub-domains and a proof of the existence of the limit of the sum mentioned above in the case of continuous functions $f(N)$ and a class of discontinuous functions.

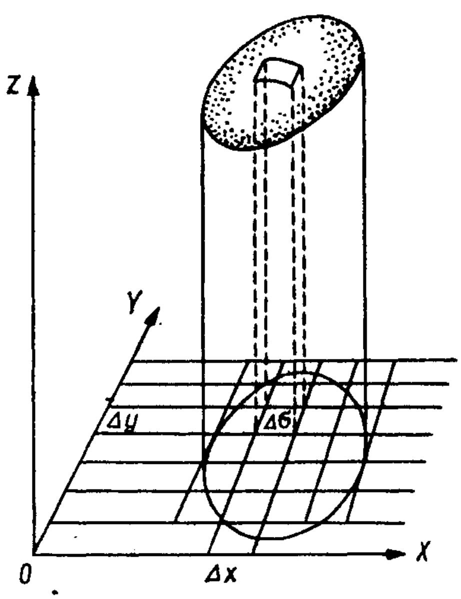

FIG. 38

56. Evaluation of double integrals. By considering a double integral as a volume, we can deduce a method for reducing a double integral to an iterated integral.

If the domain (σ) is referred to rectangular coordinates, we can let the elementary domains $\Delta\sigma$ consist of rectangles of sides Δx, Δy formed by lines parallel to the coordinate axes (Fig. 38). Let (x, y) be the coordinates of the point N. We can now write:

$$f(N) = f(x, y); \quad \Delta\sigma = \Delta x\,\Delta y; \quad \mathrm{d}\sigma = \mathrm{d}x\,\mathrm{d}y$$

and

$$\iint\limits_{(\sigma)} f(N)\,\mathrm{d}\sigma = \lim \sum_{(\sigma)} f(x, y)\,\Delta x\,\Delta y = \iint\limits_{(\sigma)} f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

On the other hand, on applying what was said in [54] about evaluating a volume in terms of an iterated integral, we can write:

$$\iint\limits_{(\sigma)} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_a^b \mathrm{d}x \int_{y_1}^{y_2} f(x, y)\,\mathrm{d}y = \int_\alpha^\beta \mathrm{d}y \int_{x_1}^{x_2} f(x, y)\,\mathrm{d}x, \qquad (7)$$

which gives us the rule for evaluating a double integral, independently of the geometrical significance of function $f(x, y)$.

If the first integration is with respect to y, x is meantime reckoned constant, whilst the limits y_1 and y_2 are the functions of x given by (2) [54]. And similarly, if the first integration is with respect to x. The limits of the first integration in the iterated integral are definite constants, independent of the variable of the second integration, only in the case when the domain of integration is a rectangle with sides parallel to the axes. If (σ) is a rectangle (Fig. 39), bounded by the lines

$$x = a; \quad x = b; \quad y = \alpha; \quad y = \beta,$$

we have

$$\iint_{(\sigma)} f(x, y)\,dx\,dy = \int_a^b dx \int_\alpha^\beta f(x, y)\,dy = \int_\alpha^\beta dy \int_a^b f(x, y)\,dx. \tag{8}$$

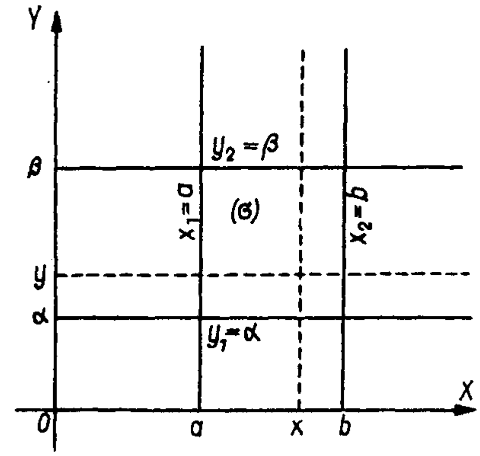

FIG. 39

The quantity $d\sigma = dx\,dy$ *is called an elementary area in rectangular coordinates.*

We remark that the first integration of (7) with respect to y with constant x corresponds to summation over the rectangles consisting of strips parallel to axis OY, all these rectangles having the same width dx, which is taken outside the sign of the first integration. The second integration, with respect to x, corresponds to adding all these sums. We give an accurate derivation of (8) and (7) in the next article.

If lines parallel to the axes intersect the boundary of (σ) more than twice, we must proceed as mentioned in [54].

We naturally assume, here and later on, that the integrals spoken of exist [95]. A sufficient condition for this is that the integrands be continuous in (σ) as far as its boundary, which we shall assume to be the case, and that the domain (σ) satisfy the condition laid down in the basic discussion of integrals in [91].

Now let (σ) be referred to polar coordinates. The equation of the surface (S) will now have to be written in the form $z = f(r, \varphi)$, where (r, φ) are the polar coordinates.

We now get the elements $\Delta\sigma$ by drawing the families of lines $r = \text{const.}$ and $\varphi = \text{const.}$, i.e. concentric circles and radius vectors passing through the origin (Fig. 40). If we take the particular $\Delta\sigma$ formed by the circles of radii r and $(r + \Delta r)$ and by the radii at angles φ and $(\varphi + \Delta\varphi)$, the curvilinear figure can be replaced to an accuracy of higher order infinitesimals by the rectangle of sides Δr and $r\,\Delta\varphi$, so that

$$\Delta\sigma = r\,\Delta r\,\Delta\varphi,$$

and we can write:

$$\iint\limits_{(\sigma)} f(N)\,\mathrm{d}\sigma = \lim \sum_{(\sigma)} f(r, \varphi)\, r\Delta r\,\Delta\varphi = \iint\limits_{(\sigma)} f(r, \varphi)\, r\,\mathrm{d}r\,\mathrm{d}\varphi.$$

The function under the double integral obtained here is $rf(r, \varphi)$. The same rule as above can be used for reducing this to an iterated integral, except that the roles of x and y are played here by r and φ.

The first integration with respect to r at constant φ corresponds to summation over the elements $\Delta\sigma$ contained between the two radii φ and $(\varphi + \mathrm{d}\varphi)$, $\mathrm{d}\varphi$ being taken outside the sign of the first integration. The second integration with respect to φ corresponds to addition of all these sums. When applying the rule, we first note the extreme values α and β of the argument φ (the extreme values of x in [54]), then the radius vectors r_1 and r_2 of, respectively, the points of entry into and departure from (σ) of the vector $\varphi = \text{const.}$ (which corresponds to finding y_1 and y_2 in [54]). Given these values, we have:

$$\iint\limits_{(\sigma)} f(N)\,\mathrm{d}\sigma = \iint\limits_{(\sigma)} f(r, \varphi)\, r\,\mathrm{d}r\,\mathrm{d}\varphi = \int\limits_{\alpha}^{\beta} \mathrm{d}\varphi \int\limits_{r_1}^{r_2} f(r, \varphi)\, r\,\mathrm{d}r, \qquad (9)$$

where r_1 and r_2 are known functions of φ.

Figure 40 illustrates the case when the origin lies outside the contour (l). If the origin is inside the contour, φ can be taken as varying from 0 to 2π and r from 0 to r_2 for a given φ, where r_2 is obtained from the equation of the curve (l): $r_2 = \psi(\varphi)$, so that we have (Fig. 41):

$$\iint\limits_{(\sigma)} f(N)\,\mathrm{d}\sigma = \int\limits_{0}^{2\pi} \mathrm{d}\varphi \int\limits_{0}^{r_2} f(r, \varphi)\, r\,\mathrm{d}r.$$

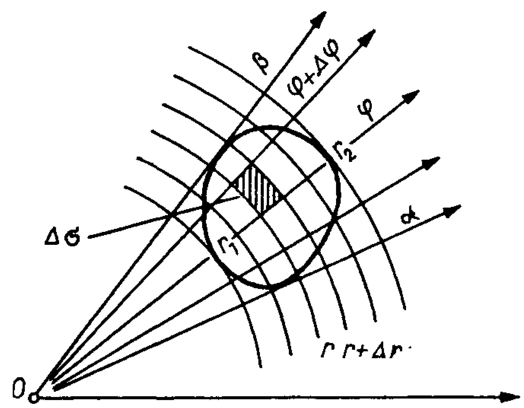

Fig. 40

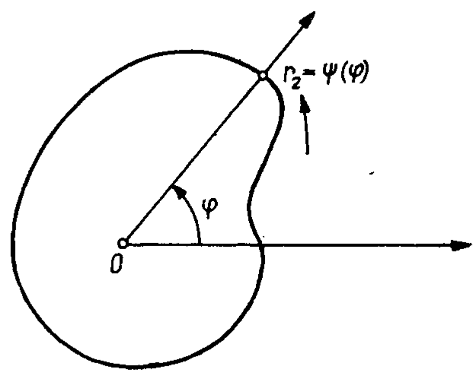

Fig. 41

The expression

$$r\,\mathrm{d}r\,\mathrm{d}\varphi \tag{10}$$

is called an *elementary area in polar coordinates.*

In particular, if $f(N) = 1$, we get the expression obtained in [I, 102] for the area under a curve in polar coordinates:

$$\int_\alpha^\beta \mathrm{d}\varphi \int_{r_1}^{r_2} r\,\mathrm{d}r = \frac{1}{2}\int_\alpha^\beta (r_2^2 - r_1^2)\,\mathrm{d}\varphi.$$

(The expression in [I, 102] is for the case when $r_2 = r$ and $r_1 = 0$.)

Example. We find the volume included between a sphere of radius a and a right circular cylinder of radius $a/2$, passing through the centre of the sphere (Fig. 42). We take the centre of the sphere as origin, the XOY plane perpendicular to the axis of the cylinder, and the axis OX from the centre of the sphere to the point of intersection of the axis of the cylinder and the XOY plane. We can say by symmetry that the required volume is four times the part of the cylinder bounded by the ZOX, XOY planes and by the upper hemisphere.

The domain of integration here is half the base of the cylinder, the contour of which consists of the semicircle

$$r = a\cos\varphi$$

with the corresponding part of axis OX; the angle φ varies from 0 to $\pi/2$, and the corresponding radius vector from the axis OX to axis OY.

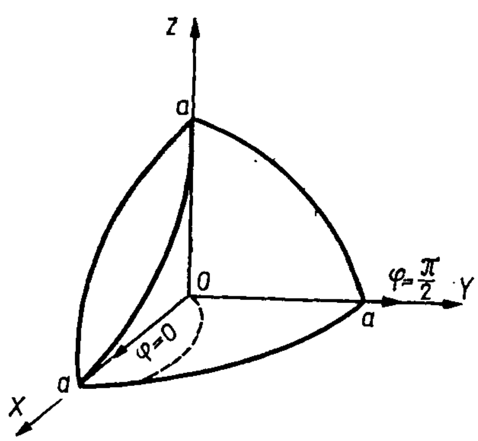

Fig. 42

The equation of the spherical surface,:

$$x^2 + y^2 + z^2 = a^2$$

can be written here as

$$z^2 = a^2 - (x^2 + y^2); \quad z = \sqrt{a^2 - r^2}.$$

The required volume is therefore

$$v = 4 \int\limits_0^{\frac{\pi}{2}} d\varphi \int\limits_0^{a\cos\varphi} \sqrt{a^2 - r^2}\, r\, dr = 4 \int\limits_0^{\frac{\pi}{2}} \left[-\frac{1}{3} (a^2 - r^2)^{\frac{3}{2}} \right] \Bigg|_{r=0}^{r=a\cos\varphi} d\varphi =$$

$$= \frac{4}{3} \int\limits_0^{\frac{\pi}{2}} (a^3 - a^3 \sin^3\varphi)\, d\varphi = \frac{4}{3} a^3 \left[\varphi + \cos\varphi - \frac{\cos^3\varphi}{3} \right] \Bigg|_{\varphi=0}^{\varphi=\frac{\pi}{2}} =$$

$$= \frac{4}{3} a^3 \left(\frac{\pi}{2} - \frac{2}{3} \right).$$

57. Curvilinear coordinates. We defined an elementary area in the previous article, and considered the problem of evaluating integrals in the case of rectangular (x, y) and polar (r, φ) coordinates. We now deal with these problems with any coordinates (u, v), where the new variables u, v that replace x, y are given by

$$\varphi(x, y) = u; \quad \psi(x, y) = v. \tag{11}$$

If we fix the value of u and take v variable, we get a family of lines on a plane. Similarly, if we fix v and take u variable, we get a second family of lines. The lines of the two families can be either curved or straight (Fig. 43).

The position of a point M on the plane is defined by a pair of numbers (x, y) or, by (11), by the pair of numbers (u, v), which are referred to as *the curvilinear coordinates of the point M*. We get expressions for the rectangular (x, y) in terms of the curvilinear (u, v) coordinates on solving equations (11) with respect to x and y:

$$x = \varphi_1(u, v); \quad y = \psi_1(u, v). \tag{12}$$

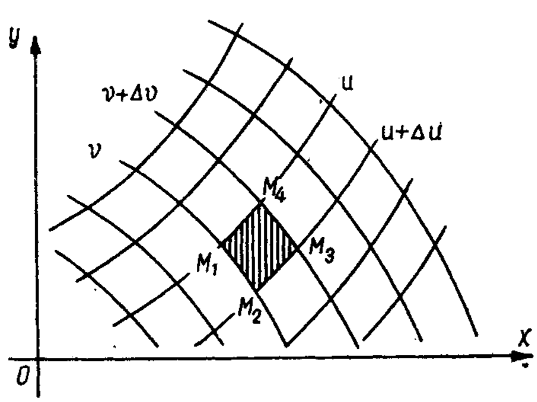

FIG. 43

In the case of polar coordinates, u becomes r and v becomes φ. The lines of the

constant u and of the constant v, mentioned above, are called coordinate lines in curvilinear coordinates (u, v); they form two families of lines (circles and radii in polar coordinates).

We now find the elementary area $d\sigma$ in curvilinear coordinates. For this, we consider the area $M_1 M_2 M_3 M_4$ (Fig. 43), formed by two pairs of infinitesimally distant coordinate lines:

$$\varphi(x, y) = u; \quad \varphi(x, y) = u + du,$$
$$\psi(x, y) = v; \quad \psi(x, y) = v + dv.$$

The coordinates of the vertices of $M_1 M_2 M_3 M_4$ are given to an accuracy of higher order infinitesimals [I, 68] by:

$$(M_1)\, x_1 = \varphi_1(u, v); \quad y_1 = \psi_1(u, v).$$

$$(M_2)\, x_2 = \varphi_1(u + du, v) = \varphi_1(u, v) + \frac{\partial \varphi_1(u, v)}{\partial u}\, du;$$

$$y_2 = \psi_1(u + du, \quad v) = \psi_1(u, v) + \frac{\partial \psi_1(u, v)}{\partial u}\, du.$$

$$(M_3)\, x_3 = \varphi_1(u + du,\ v + dv) = \varphi_1(u, v) + \frac{\partial \varphi_1(u, v)}{\partial u}\, du + \frac{\partial \varphi_1(u, v)}{\partial v}\, dv;$$

$$y_3 = \psi_1(u + du, v + dv) = \psi_1(u, v) + \frac{\partial \psi_1(u, v)}{\partial u}\, du + \frac{\partial \psi_1(u, v)}{dv}\, dv;$$

$$(M_4)\, x_4 = \varphi_1(u, v + dv) = \varphi_1(u, v) + \frac{\partial \varphi_1(u, v)}{\partial v}\, dv;$$

$$y_4 = \psi_1(u, v + dv) = \psi_1(u, v) + \frac{\partial \psi_1(u, v)}{\partial v}\, dv.$$

It follows at once from these expressions that $x_2 - x_1 = x_3 - x_4$ and $y_2 - y_1 = y_3 - y_4$, and hence the segments $M_1 M_2$ and $M_4 M_3$ are equal and have the same direction. The same can be said of $M_1 M_4$ and $M_2 M_3$, i.e. $M_1 M_2 M_3 M_4$ is a parallelogram to an accuracy of higher order infinitesimals; its area is thus twice the area of triangle $M_1 M_2 M_3$, and therefore, by a familiar expression of analytic geometry:

$$d\sigma = |x_1(y_2 - y_3) - y_1(x_2 - x_3) + (x_2 y_3 - x_3 y_2)|.$$

On substituting the expressions for the coordinates, we get the formula for an elementary area in any curvilinear coordinates:

$$d\sigma = \left| \frac{\partial \varphi_1(u, v)}{\partial u} \frac{\partial \psi_1(u, v)}{\partial v} - \frac{\partial \varphi_1(u, v)}{\partial v} \frac{\partial \psi_1(u, v)}{\partial u} \right| du\, dv = |D|\, du\, dv,$$

where D is called the *functional determinant* (or *Jacobian*) of the functions $\varphi_1(u, v)$ and $\psi_1(u, v)$ with respect to u and v:

$$D = \frac{\partial \varphi_1 (u, v)}{\partial u} \frac{\partial \psi_1 (u, v)}{\partial v} - \frac{\partial \varphi_1 (u, v)}{\partial v} \frac{\partial \psi_1 (u, v)}{\partial u}.$$

Finally, the formula for change of variables in a double integral is:

$$\iint_{(\sigma)} f(x, y)\, d\sigma = \iint_{(\sigma)} F(u, v)\, |D|\, du\, dv, \tag{13}$$

where $F(u, v)$ denotes the function of u and v to which $f(x, y)$ is transformed as a result of (12). The limits of integration with respect to u and v are found from the form of (σ) in a manner similar to that described in [56] for polar coordinates.

We took u and v in transformations (11) as the new curvilinear coordinates of a point, and reckoned the plane invariable. Conversely, we can take u and v as the previous rectangular coordinates and let (11) transform the plane so that points with rectangular coordinates (x, y) become points with rectangular coordinates (u, v). Such a transformation changes (σ) to a new domain (Σ), and (13) now has to be written as:

$$\iint_{(\sigma)} f(x, y)\, d\sigma = \iint_{(\Sigma)} F(u, v)\, |D|\, du\, dv,$$

where the u and v are the rectangular coordinates of points of the domain (Σ) and the limits of the integration over (Σ) are found in the manner of [56]. On putting $f(x, y) = F(u, v) = 1$, we get an expression for the area σ of the domain (σ) as an integral over (Σ):

$$\sigma = \iint_{(\Sigma)} |D|\, du\, dv.$$

One consequence of this is that, from our new point of view, $|D|$ at any point N of the domain (Σ) is the coefficient for the change of area at N on transition from (Σ) to (σ), i.e. it is the limit of the ratio of the area of a domain in (σ) containing the image of N to the area of the corresponding domain in (Σ) which contains N, when the latter domain contracts to the point N. We deal with the change of variables in a double integral in more detail from this point of view in [77].

Examples. **1.** We take the circle $x^2 + y^2 \leqslant 1$ of unit radius and with centre at the origin in the XY plane. We introduce new variables in accordance with the formulae for passage to polar coordinates: $x = r \cos \varphi$, $y = r \sin \varphi$ but in fact take r and φ as rectangular coordinates; thus a point with rectangular coordinates (x, y) becomes a point with rectangular coordinates (r, φ). Clearly,

with this, our circle becomes a rectangle whose sides are the straight lines $x = 0$, $x = 1$, $y = 0$, $y = 2\pi$ (or $r = 0$, $r = 1$, $\varphi = 0$, $\varphi = 2\pi$); the origin $x = y = 0$ corresponds to the complete side $r = 0$ of the rectangle, whilst the opposite sides $\varphi = 0$ and $\varphi = 2\pi$ both correspond to the same radius of the circle. If we use the rule expressed by (8) for reducing a double to an iterated integral, we can see at once that the limits for integration in polar coordinates over our circle must be $r = 0$ and $r = 1$ for r, and $\varphi = 0$ and $\varphi = 2\pi$ for φ. The rules given in [56] for finding the limits of integration in polar coordinates can be similarly explained.

In the present case

$$D = \frac{\partial(r\cos\varphi)}{\partial r} \cdot \frac{\partial(r\sin\varphi)}{\partial\varphi} - \frac{\partial(r\cos\varphi)}{\partial\varphi} \cdot \frac{\partial(r\sin\varphi)}{\partial r} = r,$$

and, as we saw above, $d\sigma = r dr d\varphi$.

2. We take a right-angled triangle (σ), formed by the coordinate axes and by the straight line $x + y = a$, as a further example of the second aspect of double integrals. The coordinates of points inside (σ) must satisfy the inequalities:

$$x > 0; \quad y > 0; \quad x + y < a. \tag{14}$$

We introduce new variables (u, v) by putting

$$x + y = u; \quad ay = uv,$$

i.e.

$$u = x + y; \quad v = \frac{ay}{x + y},$$

or

$$x = \frac{u(a - v)}{a}; \quad y = \frac{uv}{a}.$$

We take (u, v) as a new system of rectilinear rectangular coordinates. It follows from the last expression that inequalities (14) are equivalent in the new system to: $0 < u < a$, $0 < v < a$, which define a square (Σ) with a vertex at the origin and sides along the axes. Every point (x, y) of (σ) corresponds to a definite point (u, v) of (Σ), and conversely. We get for D:

$$D = \frac{a - v}{a} \cdot \frac{u}{a} - \frac{u}{a} \cdot \frac{v}{a} = \frac{u}{a},$$

and (13) takes the form:

$$\iint\limits_{(\sigma)} f(x, y)\, dx\, dy = \iint\limits_{(\Sigma)} F(u, v) \frac{u}{a}\, du\, dv,$$

or, on introducing the limits of integration in accordance with (7) and (8)

$$\int_0^a dx \int_0^{a-x} f(x, y)\, dy = \frac{1}{a} \int_0^a u\, du \int_0^a F(u, v)\, dv.$$

58. Triple integrals. Instead of interpreting the double integral of [55] as the volume of a body, we can take it as representing the mass of a material distributed over the plane domain (σ). Suppose that Δm is the amount of material on the elementary area $\Delta\sigma$ which contains a given point N. If the ratio $\Delta m/\Delta\sigma$ tends to a definite limit as $\Delta\sigma$ contracts to N, this limit $f(N)$ gives the density of the surface distribution of material at the point N:

$$\lim \frac{\Delta m}{\Delta\sigma} = f(N).$$

On dividing (σ) into small elements $\Delta\sigma$, the mass on a single element will be approximately $f(N)\,\Delta\sigma$, and we can write the approximation for the total mass on (σ):

$$m \sim \sum_{(\sigma)} f(N)\,\Delta\sigma,$$

where the summation extends over all the $\Delta\sigma$ that make up (σ). This approximation increases in accuracy as the $\Delta\sigma$ diminish. We have in the limit, on indefinite overall contraction of every element $\Delta\sigma$, and with their number indefinitely increasing:

$$m = \lim \sum_{(\sigma)} f(N)\,\Delta\sigma = \int\limits_{(\sigma)}\!\!\int f(N)\,\mathrm{d}\sigma.$$

In the same way, consideration of the mass of a spatial distribution of matter leads to the concept of a triple integral. We take a volume (v) bounded by a closed surface (S) and let matter of total mass m be distributed throughout the volume. We divide (v) into a large number n of small elements Δv, and let the mass of Δv be Δm; assuming that the ratio

$$\frac{\Delta m}{\Delta v}$$

tends to a limit on indefinite contraction of a Δv to the point M contained in it, the limit is called the *density of the (spatial) distribution at the point* M.

We denote the limit by $f(M)$:

$$\lim \frac{\Delta m}{\Delta v} = f(M).$$

We can write approximately, as above:

$$m \sim \sum_{(v)} f(M)\,\Delta v,$$

where the summation extends over all the elements Δv that make up (v).

We have in the limit, on indefinite contraction of every element Δv:

$$m = \lim \sum_{(v)} f(M) \, \Delta v.$$

The above physical example leads us to a general definition of triple integrals, analogous to that for double integrals. Let (v) be a bounded domain of three dimensional space and let $f(M)$ be a function of points of the domain, i.e. a function that takes a definite value at every point M of the domain. We divide (v) into n parts, of volumes $\Delta v_1, \Delta v_2, \ldots, \Delta v_n$ and containing respectively the points $M_1, M_2, \ldots, M_n$.

We form the sum:

$$\sum_{k=1}^{n} f(M_k) \, \Delta v_k. \tag{15}$$

The limit of this sum, on indefinite decrease of every sub-domain and indefinite increase in their number, is called the triple integral of the function $f(M)$ over the volume (v):

$$\iiint_{(v)} f(M) \, \mathrm{d}v = \lim \sum_{k=1}^{n} f(M_k) \, \Delta v_k.$$

Remark [cf. 55]. Let d_k be the maximum distance between two points of the sub-domain Δv_k (the diameter of the sub-domain) and let d be the greatest of the numbers $d_1, d_2, \ldots, d_n$. Indefinite decrease of every sub-domain means that $d \to 0$. If I denotes the magnitude of the integral, the definition given above is equivalent to the following: for any given positive number ε there exists a positive η such that

$$\left| I - \sum_{k=1}^{n} f(M_k) \, \Delta v_k \right| \leqslant \varepsilon, \quad \text{provided} \quad d \leqslant \eta.$$

The rigorous theory of triple integrals will be found with that of double integrals at the end of the chapter.

If $f(M) = 1$ throughout the domain (v), we get the volume v of the domain as:

$$v = \iiint_{(v)} \mathrm{d}v.$$

The evaluation of triple integrals requires their reduction to single or double integrals, which can then be evaluated by known methods.

We take a system of rectangular coordinates in space and suppose, for simplicity, that any line parallel to a coordinate axis cuts the surface (S) bounding the volume (v) in not more than two points. We draw a cylinder which projects the surface (S) on to the XOY plane to form the domain (σ_{xy}) (Fig. 44). The tangent curve of the cylinder and surface cuts (S) into two parts:

$$\text{(I)} \quad z_1 = \varphi_1(x, y);$$
$$\text{(II)} \quad z_2 = \varphi_2(x, y).$$

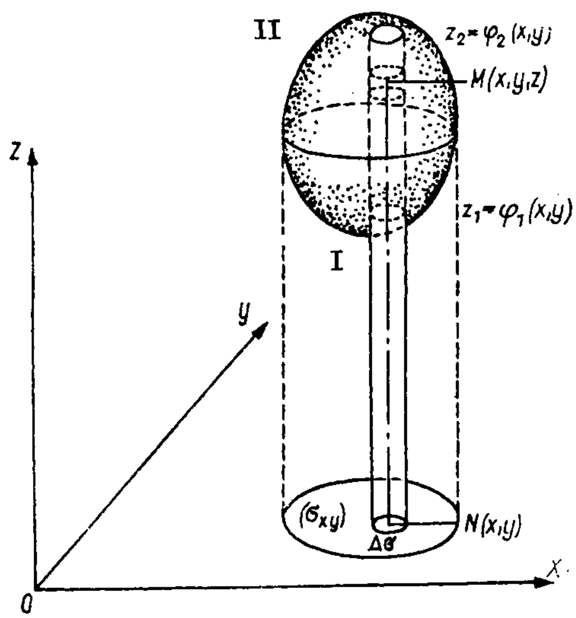

Fig. 44

A line parallel to axis OZ and passing through any point of (σ_{xy}) enters the volume (v) through section (I) and leaves it via section (II); the ordinates of the points of entry and exit, z_1 and z_2, will be known functions of (x, y).

We now agree to divide volume (v) into elements Δv as follows: we divide the area (σ_{xy}) into a large number of small elements $\Delta\sigma$; taking an element as a base, we draw a cylinder for every element; the intersection of the cylinder with volume (v) is then divided into elementary cylinders of height Δz by means of planes parallel to XOY and at a distance Δz apart. The elements of volume obtained will be given by

$$\Delta v = \Delta\sigma \, \Delta z.$$

We take the element $\Delta\sigma$ containing the point $N(x, y)$ and produce a line through it parallel to axis OZ to cut (S) in points with ordinates z_1 and z_2; then we take points $M(x, y, z)$ on this line, lying inside the corresponding elements Δv.

The sum of (15) can be written as:

$$\sum_{(v)} f(x, y, z)\,\Delta v = \sum_{(\sigma)} \Delta\sigma \sum_{(z)} f(x, y, z)\,\Delta z.$$

Having fixed $\Delta\sigma$ for the present, we let the Δz diminish, and get by the concept of definite integral:

$$\lim \sum_{(z)} f(x, y, z)\,\Delta z = \int_{z_1}^{z_2} (f(x, y, z)\,\mathrm{d}z,$$

where the x and y must be looked on as constant parameters. Hence we have approximately:

$$\sum_{(z)} f(x, y, z)\,\Delta z \sim \int_{z_1}^{z_2} f(x, y, z)\,\mathrm{d}z = \Phi(x, y).$$

But now it is clear, by the definition of double integral, that

$$\sum_{(v)} f(x, y, z)\,\Delta v \sim \sum_{(\sigma_{xy})} \Delta\sigma\,\Phi(x, y) \to \iint_{(\sigma_{xy})} \Phi(x, y)\,\mathrm{d}\sigma,$$

i.e.

$$\iiint_{(v)} f(x, y, z)\,\mathrm{d}v = \iint_{(\sigma_{xy})} \mathrm{d}\sigma \int_{z_1}^{z_2} f(x, y, z)\,\mathrm{d}z. \tag{16}$$

On setting aside the geometrical interpretation, the above arguments lead us to the following rule for evaluating triple integrals.

To reduce the triple integral

$$\iiint_{(v)} f(x, y, z)\,\mathrm{d}v$$

to a single and double integral: (1) *project the surface* (S) *bounding the volume* (v) *on the* XY *plane to form the domain* (σ_{xy}); (2) *find the coordinates* z_1 *and* z_2 *of the points of entry and exit of a line parallel to* OZ *through the point* (x, y) *of* (σ_{xy}); (3) *evaluate the integral*

$$\int_{z_1}^{z_2} f(x, y, z)\,\mathrm{d}z,$$

taking x, y constant, then evaluate the double integral

$$\iint\limits_{(\sigma_{xy})} d\sigma \int\limits_{z_1}^{z_2} f(x, y, z)\, dz.$$

The double integral may be in turn reduced to an interated integral, on using rectangular coordinates (x, y), and we finally get:

$$\iiint\limits_{(v)} f(x, y, z)\, dv = \int\limits_{a}^{b} dx \int\limits_{y_1}^{y_2} dy \int\limits_{z_1}^{z_2} f(x, y, z)\, dz, \tag{17}$$

where the limits (y_1, y_2) and (a, b) are found as in [54].

We suggest that the reader work out the alternative methods of reducing a triple to an iterated integral, by projecting (S) on to the YZ plane to form an area (σ_{yz}), or on to the XZ plane to form the area (σ_{xz}). The reader might also consider the more difficult cases, when lines parallel to the coordinate axes cut the surface in more than two points.

Equation (17) may be written as:

$$\iiint\limits_{(v)} f(x, y, z)\, dx\, dy\, dz = \int\limits_{a}^{b} dx \int\limits_{y_1}^{y_2} dy \int\limits_{z_1}^{z_2} f(x, y, z)\, dz.$$

The product $dx\, dy\, dz$ *is called an elementary volume in rectangular coordinates, and is obtained by dividing the volume* (v) *into infinitely small rectangular parallelepipeds by means of planes parallel to the coordinate axes.*

Formula (17) is derived rigorously at the end of the chapter. We remark that if lines parallel to the axes cut (S) in more than two points, (v) has to be sub-divided so that each part of it is not intersected at more than two points. Integration is carried out as above for each sub-division and the values added to give the integral over the whole volume (v).

If (v) is a rectangular parallelepiped, bounded by the planes

$$x = a; \quad x = b; \quad y = a_1; \quad y = b_1; \quad z = a_2; \quad z = b_2,$$

parallel to the axes, the limits for all three integrals become constants:

$$\iiint\limits_{(v)} f(x, y, z)\, dx\, dy\, dz = \int\limits_{a}^{b} dx \int\limits_{a_1}^{b_1} dy \int\limits_{a_2}^{b_2} f(x, y, z)\, dz. \tag{18}$$

59. Cylindrical and spherical coordinates. A system of rectangular coordinates in space is often not the most convenient; of the alter-

native systems, the most important are those of *cylindrical and spherical coordinates*. The position of a point in the usual rectangular system is determined by three coordinates (a, b, c), the point being situated at the intersection of the three planes $x = a$, $y = b$, $z = c$, parallel to the coordinate planes. We can think of the space here as filled by three families of mutually perpendicular planes

$$x = C_1; \quad y = C_2; \quad z = C_3,$$

where C_1, C_2, C_3 are constants, every point of space being the intersection of three planes of the families. Let the coordinates x and y now be replaced by r and φ, whilst the coordinate z is retained, where

$$x = r\cos\varphi; \quad y = r\sin\varphi; \quad z = z.$$

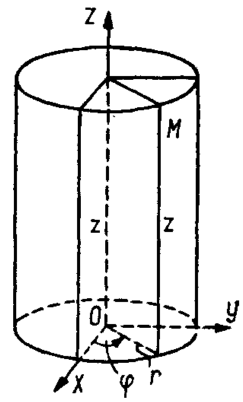

Fig. 45

The coordinate r is the distance of a point M from axis OZ and φ is the angle between the plane passing through OZ and M and the plane XOZ (Fig. 45); φ can vary from 0 to 2π, and r from 0 to $(+\infty)$. The coordinates (r, φ, z) are called the *cylindrical coordinates* of the point M. Points on axis OZ correspond to $r = 0$, their coordinate φ being indeterminate.

We have the following three coordinates surfaces in this case

$$r = C_1; \quad \varphi = C_2; \quad z = C_3.$$

The family $r = C_1$ is a family of circular cylinders with OZ as their common axis of revolution. The second family $\varphi = C_2$ is a family of half-planes passing through axis OZ, and lastly, $z = C_3$ is a family of planes parallel to the XOY plane.

On giving the variables r, φ, z increments $\Delta r, \Delta\varphi, \Delta z$, we get two neighbouring surfaces for each family, which together form an elementary volume in cylindrical coordinates. Only one coordinate varies along each edge of the element, whilst every pair of edges is orthogonal (Fig. 46). To an accuracy of higher order infinitesimals, the element can be taken as a rectangular parallelepiped with edges

$$\Delta r, \; r\Delta\varphi, \; \Delta z,$$

which leads us to the following expression for an elementary volume in cylindrical coordinates:

$$dv = r\,dr\,d\varphi\,dz$$

as well as to the expression for a triple integral in these coordinates:

$$\iiint\limits_{(v)} f(M)\,\mathrm{d}v = \iiint\limits_{(v)} f(r, \varphi, z)\, r\,\mathrm{d}r\,\mathrm{d}\varphi\,\mathrm{d}z\,, \tag{19}$$

the limits of integration being found by similar methods to those used for rectangular coordinates.

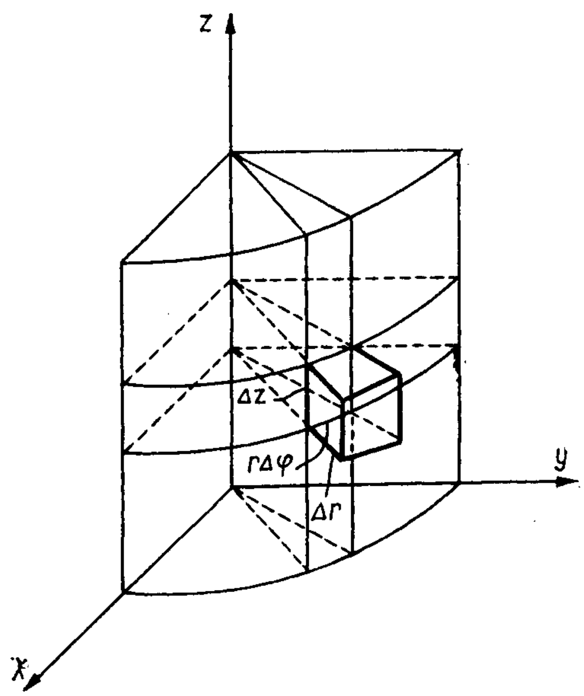

FIG. 46

Example. To find the mass of the segment of a sphere, filled with non-homogenous material whose density is proportional to the distance from the base of the segment (Fig. 47).

We locate the origin at the centre of the sphere, and take for the XOY plane the diametral plane parallel to the base of the segment; we direct axis OZ from the origin towards the segment. Let the radius of the sphere be a, the height of the segment h, and the radius of base of the segment r_0.

The equation of the sphere becomes in cylindrical coordinates:

$$r^2 + z^2 = a^2 \quad \text{or} \quad z^2 = a^2 - r^2.$$

From what was given, we can write the density as

$$f(r, \varphi, z) = b + cz,$$

where b and c are constants.

Application of (19) gives:

$$m = \iiint\limits_{(v)} (b + cz)\, r\, \mathrm{d}r\, \mathrm{d}\varphi\, \mathrm{d}z = \int\limits_0^{2\pi} \mathrm{d}\varphi \int\limits_0^{r_0} r\, \mathrm{d}r \int\limits_{a-h}^{\sqrt{a^2-r^2}} (b + cz)\, \mathrm{d}z =$$

$$= 2\pi \int\limits_0^{r_0} \left[bz + \frac{c}{2}\, z^2 \right]_{z=a-h}^{z=\sqrt{a^2-r^2}} r\, \mathrm{d}r\,.$$

We leave it to the reader to carry out the substitutions for z and the integration; the result is

$$m = bv + c\pi \frac{r_0^4}{4}\,,$$

where v is the volume of the segment.

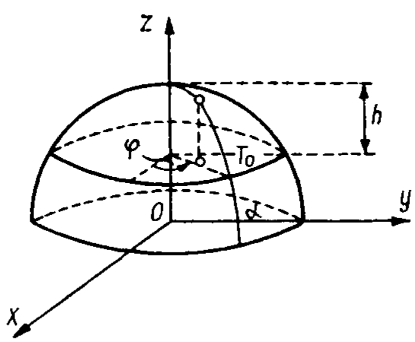

FIG. 47

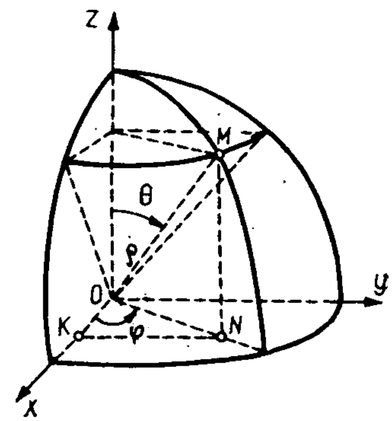

FIG. 48

We now discuss *spherical coordinates*, or *polar coordinates in space*, as they are sometimes called. Let M be a given point in space, and $\overline{OM}$ the vector drawn from the origin O to M. The position of M can be defined by the following three quantities: the length ϱ of $\overline{OM}$; the angle φ that the half-plane passing through axis OZ and M makes with the XZ plane; the angle θ that $\overline{OM}$ makes with the positive direction of OZ (Fig. 48). With this, ϱ can vary from 0 to $(+\infty)$; φ is reckoned counter-clockwise from axis OX and can vary from 0 to 2π; finally, θ is reckoned from the positive direction of OZ and can vary from 0 to π. Definite coordinates ϱ, φ, θ correspond to every point M, and conversely. We drop a perpendicular from M to the XY plane to meet it in N, then drop another perpendicular NK from N to axis OX. The rectangular coordinates x, y, z of M

are clearly given by $\overline{OK}$, $\overline{KN}$, $\overline{NM}$. We have from the right-angled triangle ONM:

$$\overline{ON} = \varrho \sin\theta$$

and on also using the right-angled triangle ONK, we finally get *the formulae for passing from rectangular to spherical coordinates:*

$$x = \varrho \sin\theta \cos\varphi, \quad y = \varrho \sin\theta \sin\varphi, \quad z = \varrho \cos\theta .$$

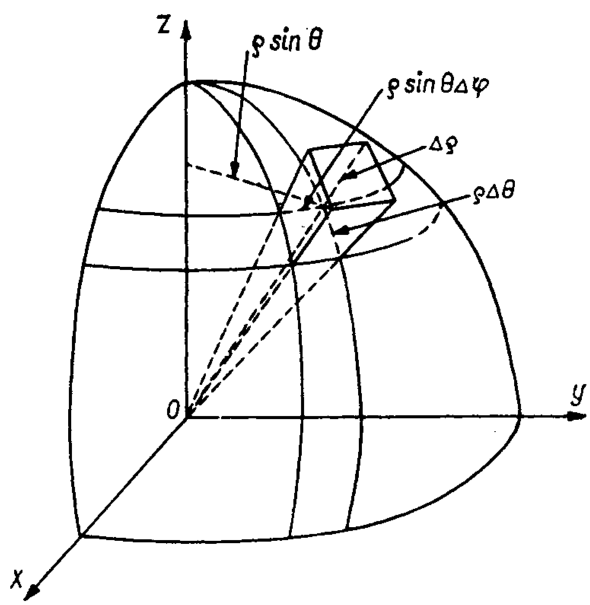

Fig. 49

We consider the families of coordinate surfaces:

$$\varrho = c_1 \quad \varphi = c_2, \quad \theta = c_3 .$$

The first is evidently a family of spheres with centres at the origin; the second is a family of half-planes passing through axis OZ, whilst the third is a family of circular cones with axis of revolution on OZ. We remark that the origin O corresponds to $\varrho = 0$, the values of φ and θ being indeterminate. The coordinate φ is indeterminate for all points lying on OZ, where $\theta = 0$ or π.

We get an elementary volume in spherical coordinates by giving variables ϱ, φ and θ infinitesimal increments $\Delta\varrho$, $\Delta\varphi$ and $\Delta\theta$. Only one coordinate varies along each edge of the element, and pairs of adjacent edges are orthogonal (Fig. 49). The element can be

looked on, to an accuracy of higher order infinitesimals, as a rectangular parallelepiped of sides

$$d\varrho,\ \varrho\, d\theta,\ \varrho \sin\theta\, d\varphi,$$

so that the elementary volume is given by

$$dv = \varrho^2 \sin\theta\, d\varrho\, d\theta\, d\varphi,$$

whence we can write a triple integral in spherical coordinates as:

$$\iiint\limits_{(v)} f(M)\, dv = \iiint\limits_{(v)} f(\varrho, \theta, \varphi)\, \varrho^2 \sin\theta\, d\varrho\, d\theta\, d\varphi. \tag{20}$$

Reduction of the triple integral to an iterated integral can be carried out as follows, for instance: we take the central projection (σ) of the volume (v) from the origin on to the sphere of unit radius (Fig. 50) [if the origin lies inside (v), (σ) is the entire surface of the unit sphere]. We draw radius vectors passing through points of (σ), which in the simplest case enter (v) at a point of radius vector ϱ_1 and leave at a second point of radius vector ϱ_2 [we put $\varrho_1 = 0$ when the origin lies in (v)]. Now we have:

$$\iiint\limits_{(v)} f(\varrho, \theta, \varphi)\, \varrho^2 \sin\theta\, d\varrho\, d\theta\, d\varphi = \iint\limits_{(\sigma)} \sin\theta\, d\theta\, d\varphi \int\limits_{\varrho_1}^{\varrho_2} f(\varrho, \theta, \varphi)\, \varrho^2\, d\varrho,$$

where ϱ_1 and ϱ_2 are known functions of θ and φ. The limits of integration for θ and φ are determined by the shape of (σ).

Example. To find the mass of a sphere made up of concentric layers of different densities. We can take the density here as depending on ϱ only and equal to $f(\varrho)$, so that the mass is

$$m = \iiint\limits_{(v)} f(\varrho)\, \varrho^2 \sin\theta\, d\varrho\, d\theta\, d\varphi = \int\limits_0^{2\pi} d\varphi \int\limits_0^{\pi} \sin\theta\, d\theta \int\limits_0^{a} f(\varrho)\, \varrho^2\, d\varrho = 4\pi \int\limits_0^{a} f(\varrho)\, \varrho^2\, d\varrho.$$

If the density is constant and equal to unity, we get the expression for the volume of a sphere:

$$v = 4\pi \int\limits_0^{a} \varrho^2\, d\varrho = \frac{4\pi a^3}{3}.$$

Remark. An important geometric significance attaches to $\sin\theta\, d\theta\, d\varphi$, this being the elementary area into which the surface of the unit

sphere is divided by meridians and parallel circles (Fig. 51). If we divide the surface of the unit sphere into elements $d\sigma$ of arbitrary shape, we have:†

$$\iiint_{(v)} f(M)\,dv = \iint_{(\sigma)} d\sigma \int_{\varrho_1}^{\varrho_2} f(M)\,\sigma^2\,d\varrho,$$

where (σ) is the domain obtained on projecting centrally from the origin the volume concerned on to the surface of the sphere.

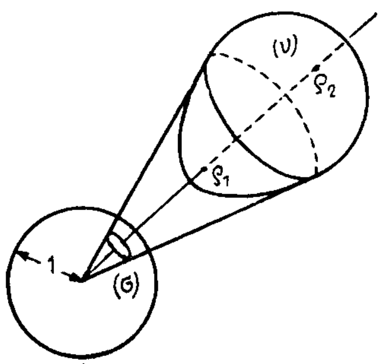

FIG. 50

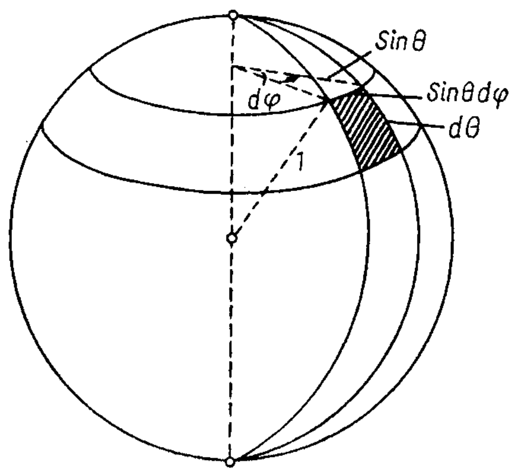

FIG. 51

Let an elementary cone be drawn, with its vertex at the centre of the unit sphere and having the contour of the element $d\sigma$ as directrix. *The area* $d\sigma$ *measures the solid angle subtended at the centre by the element of any surface* (S) *cut out from* (S) *by this elementary cone.*

60. Curvilinear coordinates in space. The position of a point in space is determined in the general case of curvilinear coordinates by three quantities q_1, q_2, q_3, related to the rectangular coordinates x, y, z by the formulae:

$$\varphi(x, y, z) = q_1; \quad \psi(x, y, z) = q_2; \quad \omega(x, y, z) = q_3. \tag{21}$$

The three families of coordinate surfaces are obtained on putting q_1, q_2, q_3 equal to constants. Three pairs of indefinitely adjacent coordinate surfaces form an elementary volume dv. We omit the proof and

† $d\sigma$ denotes either the element itself or its area.

simply give the result, analogous to that obtained in [57] for two dimensions. The elementary volume $\mathrm{d}v$ can be considered as a parallelepiped, to an accuracy of higher order infinitesimals; on solving (21) for x, y, z, we get:

$$x = \varphi_1(q_1, q_2, q_3)\,; \quad y = \psi_1(q_1, q_2, q_3); \quad z = \omega_1(q_1, q_2, q_3), \qquad (21_1)$$

and $\mathrm{d}v$ will be given by:

$$\mathrm{d}v = |D|\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3\,,$$

so that the formula for change of variables in the triple integral is:

$$\iiint\limits_{(v)} f(x, y, z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \iiint\limits_{(v)} F(q_1, q_2, q_3)\,|D|\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}q_3,$$

where $F(q_1, q_2, q_3)$ is derived from $f(x, y, z)$ by transformations (21_1) and D is the functional determinant of x, y, z with respect to q_1, q_2, q_3:

$$D = \frac{\partial\varphi_1}{\partial q_1}\left(\frac{\partial\psi_1}{\partial q_1}\cdot\frac{\partial\omega_1}{\partial q_2} - \frac{\partial\psi_1}{\partial q_3}\cdot\frac{\partial\omega_1}{\partial q_2}\right) + \frac{\partial\varphi_1}{\partial q_2}\left(\frac{\partial\psi_1}{\partial q_3}\cdot\frac{\partial\omega_1}{\partial q_1} - \frac{\partial\psi_1}{\partial q_1}\cdot\frac{\partial\omega_1}{\partial q_3}\right) + \\ + \frac{\partial\varphi_1}{\partial q_3}\left(\frac{\partial\psi_1}{\partial q_1}\cdot\frac{\partial\omega_1}{\partial q_2} - \frac{\partial\psi_1}{\partial q_2}\cdot\frac{\partial\omega_1}{\partial q_1}\right).$$

As in [57], the transformations (21) can be looked on alternatively as a deformation of space, so that a point with rectangular coordinates (x, y, z) becomes a point with rectangular coordinates (q_1, q_2, q_3), D being interpreted now as the coefficient of the change of volume at a given point on passing from (q_1, q_2, q_3) to (x, y, z).

The reduction of a triple integral in coordinates q_1, q_2, q_3 to three single integrals, and the determination of the corresponding limits, follow the same lines as for double integrals [57].

The reader familiar with determinants may notice that D can be written as the following third order determinant:

$$D = \begin{vmatrix} \dfrac{\partial\varphi_1}{\partial q_1}, & \dfrac{\partial\psi_1}{\partial q_1}, & \dfrac{\partial\omega_1}{\partial q_1} \\ \dfrac{\partial\varphi_1}{\partial q_2}, & \dfrac{\partial\psi_1}{\partial q_2}, & \dfrac{\partial\omega_1}{\partial q_2} \\ \dfrac{\partial\varphi_1}{\partial q_3}, & \dfrac{\partial\psi_1}{\partial q_3}, & \dfrac{\partial\omega_1}{\partial q_3} \end{vmatrix}.$$

A detailed treatment of these determinants appears in Volume III.

Example. Let (v) be a tetrahedron, bounded by $x + y + z = a$ and by the coordinate planes, and defined by the inequalities:

$$x > 0; \quad y > 0; \quad z > 0; \quad x + y + z < a.$$

We introduce new variables:

$$x + y + z = q_1; \quad a(y + z) = q_1 q_2; \quad a^2 z = q_1 q_2 q_3$$

and interpret (q_1, q_2, q_3) as rectilinear rectangular coordinates. It follows from the above expressions that:

$$q_1 = x + y + z; \quad q_2 = \frac{a(y+z)}{x+y+z}; \qquad q_3 = \frac{az}{y+z}$$

or

$$x = \frac{q_1(a - q_2)}{a}; \quad y = \frac{q_1 q_2 (a - q_3)}{a^2}; \quad z = \frac{q_1 q_2 q_3}{a^2}.$$

It is easily seen, after the manner of [57], that the tetrahedron (v) becomes the cube (v_1) $0 < q_1 < a$; $0 < q_2 < a$; $0 < q_3 < a$. We find that here, $D = q_1^2 q_2 / a^3$, so that the transformation is expressed by

$$\iiint\limits_{(v)} f(x, y, z)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z = \iiint\limits_{(v_1)} F(q_1, q_2, q_3) \frac{1}{a^3} q_1^2 q_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}q_3;$$

or, on finding the limits of integration:

$$\int_0^a \mathrm{d}x \int_0^{a-x} \mathrm{d}y \int_0^{a-x-y} f(x, y, z)\, \mathrm{d}z = \frac{1}{a^3} \int_0^a q_1^2\, \mathrm{d}q_1 \int_0^a q_2\, \mathrm{d}q_2 \int_0^a F(q_1, q_2, q_3)\, \mathrm{d}q_3.$$

61. Basic properties of multiple integrals. We demonstrated earlier the basic properties of definite integrals by making direct use of their definition as the limit of a sum [I, 94]. The same approach can be used for the properties of multiple integrals. We shall assume for simplicity that all the functions are continuous, so that the integrals have a meaning unconditionally.

I. A constant factor can be taken outside the integral sign:

$$\iint\limits_{(\sigma)} af(N)\, \mathrm{d}\sigma = a \iint\limits_{(\sigma)} f(N)\, \mathrm{d}\sigma.$$

II. The integral of an algebraic sum is equal to the sum of the integrals of the terms:

$$\iint\limits_{(\sigma)} [f(N) - \varphi(N)]\, \mathrm{d}\sigma = \iint\limits_{(\sigma)} f(N)\, \mathrm{d}\sigma - \iint\limits_{(\sigma)} \varphi(N)\, \mathrm{d}\sigma.$$

III. If the domain (σ) is divided into a finite number of parts, say into two parts (σ_1) and (σ_2), the integral over the whole domain is equal to the sum of the integrals over the parts:

$$\iint\limits_{(\sigma)} f(N)\, \mathrm{d}\sigma = \iint\limits_{(\sigma_1)} f(N)\, \mathrm{d}\sigma + \iint\limits_{(\sigma_2)} f(N)\, \mathrm{d}\sigma.$$

IV. If $f(N) \leqslant \varphi(N)$ in (σ),

$$\iint\limits_{(\sigma)} f(N)\,\mathrm{d}\sigma \leqslant \iint\limits_{(\sigma)} \varphi(N)\,\mathrm{d}\sigma.$$

In particular:

$$\left|\iint\limits_{(\sigma)} f(N)\,\mathrm{d}\sigma\right| \leqslant \iint\limits_{(\sigma)} |f(N)|\,\mathrm{d}\sigma.$$

V. If $\varphi(N)$ preserves the same sign in (σ), the mean value theorem is valid, i.e.

$$\iint\limits_{(\sigma)} f(N)\,\varphi(N)\,\mathrm{d}\sigma = f(N_0)\iint\limits_{(\sigma)} \varphi(N)\,\mathrm{d}\sigma,$$

where N_0 is a point lying inside (σ).

In particular, we have with $\varphi(N) = 1$:

$$\iint\limits_{(\sigma)} f(N)\,\mathrm{d}\sigma = f(N_0)\,\sigma,$$

where σ is the area of the domain (σ).

Triple integrals have analogous properties.

We remark that it has always been assumed, in the definition of double and triple integrals as limits of a sum, that the domain of integration is finite, and that the integrand $f(N)$ is bounded throughout the domain, i.e. there exists a number A, such that $|f(N)| < A$ at every point N of the domain of integration. If these conditions are not fulfilled, the integral may exist as an improper integral, in a manner similar to that described for single definite integrals in [I, 97, 98]. We deal with improper multiple integrals in § 8 [86].

62. Surface areas. We take

$$z = f(x, y) \tag{22}$$

as the equation of a given surface (S) and use the notation:

$$\frac{\partial f}{\partial x} = p; \quad \frac{\partial f}{\partial y} = q. \tag{23}$$

We have seen [I, 160] that the direction-cosines of the normal (n) to (S) at the point (x, y, z) are proportional to p, q and (-1), i.e. using analytic geometry, they are given by:

$$\left.\begin{aligned} &\cos(n, X) = \frac{p}{\pm\sqrt{1+p^2+q^2}}; \quad \cos(n, Y) = \frac{q}{\pm\sqrt{1+p^2+q^2}};\\ &\cos(n, Z) = \frac{1}{\pm\sqrt{1+p^2+q^2}}. \end{aligned}\right\} \tag{24}$$

We find the area of the portion of (S) which is cut out by the cylinder (C), which projects the portion on to the XY plane as the domain (σ) (Fig. 52). We divide (σ) into small elements $\Delta\sigma$; the cylinders with bases $\Delta\sigma$ divide (S) into elements ΔS.

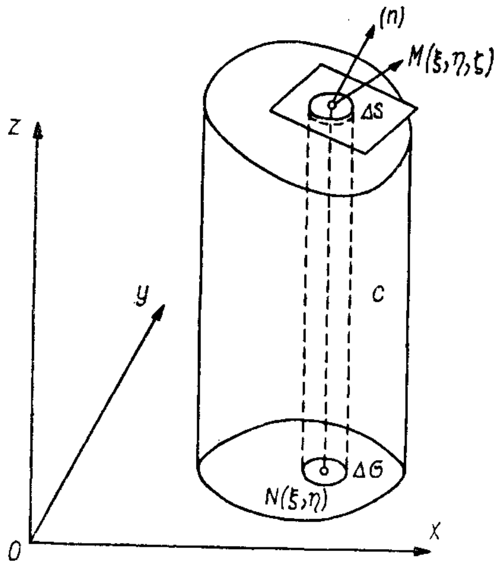

Fig. 52

Let $N(\xi, \eta)$ be a point of $\Delta\sigma$, the corresponding point of (S) being $M(\xi, \eta, \zeta)$, where $\zeta = f(\xi, \eta)$. We draw the tangent plane and normal (n) to the surface at M, and let $\Delta S'$ be the elementary area cut from the tangent plane by the cylinder of base $\Delta\sigma$.

We define the area of our portion of surface (S) as the limit of the sum of the elementary plane areas $\Delta S'$, when the number of elements $\Delta\sigma$ increases indefinitely, whilst each contracts indefinitely. We show that this limit is given by a double integral over (σ). The element $\Delta\sigma$ is the projection of the plane element $\Delta S'$ on the XY plane, whilst the normals to these elements form the angle (n, Z), the cosine of which is given by the third of expressions (24); hence†

$$\Delta\sigma = \Delta S' \frac{1}{\sqrt{1 + p^2 + q^2}} \qquad \Delta S' = \sqrt{1 + p^2 + q^2}\, \Delta\sigma,$$

† Let (S_2) be the projection of the plane domain (S_1) and let φ be the acute angle between these two planes, i.e. the angle between their normals. It is easily seen that the relationship holds: $S_2 = S_1 \cos\varphi$, where S_1 and S_2 are the areas of (S_1) and (S_2).

We show this by a rectangular dissection of the areas by two families of straight lines, one family being parallel to the line of intersection of planes (S_2) and (S_1). Let the rectangles of (S_2) be the projections of those of (S_1). A length parallel to the intersection of (S_2) and (S_1) is unchanged on projection, whilst a length in the perpendicular direction is multiplied by $\cos\varphi$; thus if $dxdy$ is an elementary rectangle of (S_1) the corresponding elementary rectangle of (S_2) will be $\cos\varphi\, dxdy$, and hence:

$$S_2 = \iint\limits_{(S_1)} \cos\varphi\, dx\, dy = \cos\varphi \iint\limits_{(S_1)} dx\, dy = S_1 \cos\varphi.$$

so that the area S of our surface becomes by definition:

$$S = \lim \sum \Delta S' = \lim \sum_{(\sigma)} \sqrt{1 + p^2 + q^2}\, \Delta\sigma.$$

The limit on the right-hand side is given by the double integral over (σ), so that

$$S = \iint_{(\sigma)} \sqrt{1 + p^2 + q^2}\, \mathrm{d}\sigma = \iint_{(\sigma)} \sqrt{1 + p^2 + q^2}\, \mathrm{d}x\, \mathrm{d}y \tag{25}$$

which is the required formula for the area of the portion of a curved surface cut out by the cylinder with generators parallel to axis OZ.

The integrand in (25) represents an elementary area $\mathrm{d}S$ of the surface. We can write, on using the expression for $\cos(n, Z)$:

$$\mathrm{d}S = \sqrt{1 + p^2 + q^2}\, \mathrm{d}\sigma_{xy} = \frac{\mathrm{d}\sigma_{xy}}{|\cos(n, Z)|} \text{ or } \mathrm{d}\sigma_{xy} = |\cos(n, Z)|\, \mathrm{d}S. \tag{26}$$

Here, $\mathrm{d}\sigma_{xy}$ is the projection of $\mathrm{d}S$ on the XY plane. We have to take the absolute value of $\cos(n, Z)$, since the elementary areas $\mathrm{d}\sigma_{xy}$ and $\mathrm{d}S$ are reckoned positive.

We assume that the p and q defined by (23) are continuous functions of (x, y). The above argument leads to the expression of the limit of the sum of areas $\Delta S'$ as an integral (25) of a continuous function, and proves, moreover, that this limit exists. A defect of the above definition of surface area is that it contains the operation of projection, which is bound up with the choice of XY plane. It can be shown that the magnitude of the surface area is in fact independent of the choice of XY plane. We also remark that, if the lines parallel to OZ meet (S) in more than one point, evaluation of the surface area by (25) requires division of the surface into separate sections and evaluation individually of the area of each.

A new definition can be given that is independent of the choice of axes. Let (S) be a piece of a smooth surface, bounded by a smooth contour. Let (S) be divided into portions $(S_1), (S_2), \ldots, (S_n)$; we take an arbitrary point M_k of each (S_k) and let p_k be the area of the projection of (S_k) on to the tangent plane to (S) at M_k. It can be shown that, given the assumptions regarding the smoothness of (S) and its contour, the sum $p_1 + p_2 + \ldots + p_n$ tends to a definite limit S if the greatest of the diameters δ of the portions tends to zero [cf. 55]. With the explicit equation (22) of the surface and with continuous

derivatives (23), this definition of surface area also leads to formula (25) for the area S. (G. M. Fikhtengol'ts, *Kurs differentsialnogo i integralnogo ischisleniya*, Vol. III, §§ 599—601).

Examples. **1.** To find the area of the portion of spherical surface considered in the example of (56).

We have

$$z = \sqrt{a^2 - x^2 - y^2}; \qquad p = \frac{-x}{\sqrt{a^2 - x^2 - y^2}} = -\frac{x}{z};$$

$$q = \frac{-y}{\sqrt{a^2 - x^2 - y^2}} = -\frac{y}{z}$$

$$\sqrt{1 + p^2 + q^2} = \sqrt{1 + \frac{x^2}{z^2} + \frac{y^2}{z^2}} = \frac{\sqrt{x^2 + y^2 + z^2}}{z} = \frac{a}{z}.$$

$$S = \iint_{(\delta)} \frac{a}{z} r \, dr \, d\varphi = 2a \int_0^{\frac{\pi}{2}} d\varphi \int_0^{a\cos\varphi} \frac{r \, dr}{\sqrt{a^2 - r^2}} =$$

$$= 2a \int_0^{\frac{\pi}{2}} \left(-\sqrt{a^2 - r^2}\right)\Bigg|_{r=0}^{r=a\cos\varphi} d\varphi = 2a^2 \int_0^{\frac{\pi}{2}} (1 - \sin\varphi) \, d\varphi = 2a^2 \left(\frac{\pi}{2} - 1\right).$$

2. To find the area of the intersection of the cylinders (Fig. 53)

$$x^2 + y^2 = a^2, \tag{27}$$

and

$$y^2 + z^2 = a^2. \tag{28}$$

It is more convenient to take y and z as the independent variables in this problem, and x as a function of these, given by (27). Integration is over the circular area in the YZ plane, with circumference given by (28). The shaded area in the figure is evidently 1/8 of the total required area, so that we have:

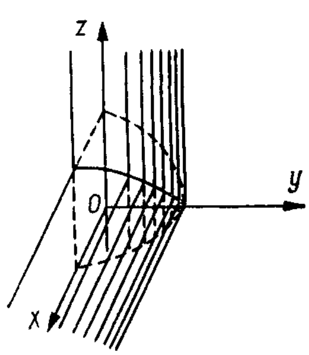

FIG. 53

$$S = 8 \iint_{(\sigma)} \sqrt{1 + p^2 + q^2} \, dy \, dz,$$

where

$$p = \frac{dx}{dy} = -\frac{y}{x}; \qquad q = \frac{dx}{dz} = 0;$$

$$\sqrt{1 + p^2 + q^2} = \frac{\sqrt{x^2 + y^2}}{x} = \frac{a}{x} = \frac{a}{\sqrt{a^2 - y^2}}$$

and hence:

$$S = 8a \int_0^a dz \int_0^{\sqrt{a^2-z^2}} \frac{dy}{\sqrt{a^2-y^2}} = 8a \int_0^a \arcsin \frac{\sqrt{a^2-z^2}}{a} dz =$$

$$= 8a \left[z \arcsin \frac{\sqrt{a^2-z^2}}{a} \Bigg|_{z=0}^{z=a} + \int_0^a \frac{z}{\sqrt{a^2-z^2}} \right] =$$

$$= -8a\sqrt{a^2-z^2} \Bigg|_{z=0}^{z=a} = 8a^2.$$

63. Integrals over a surface and Ostrogradskii's formula. The concept of a double integral over a plane domain may easily be generalized to the case of integration over a surface (curved). *Let (S) be a surface (closed or open) and $F(M)$ be a continuous function of points of the surface. We divide (S) into n parts and let $\Delta S_1, \Delta S_2, \ldots, \Delta S_n$ be the areas of the parts and $M_1, M_2, \ldots, M_n$ arbitrary points lying on them. We form the sum*

$$\sum_{k=1}^{n} F(M_k) \cdot \Delta S_k .$$

The limit of this sum on indefinite increase in the number of parts n and indefinite contraction of each part ΔS_k is called the integral of the function $F(M)$ over the surface (S):

$$\iint_{(S)} F(M)\, dS = \lim_{n\to\infty} \sum_{k=1}^{n} F(M_k)\, \Delta S_k .$$

Let a line parallel to the z axis have only one intersection with the surface (Fig. 52) and let (σ) be the projection of (S) on the plane XY. We can reduce the integral over (S) to an integral over the plane domain (σ_{xy}) by using (26) for the relationship between an elementary area of (S) and of the corresponding projection (σ_{xy}):

$$\iint_{(S)} F(M)\, dS = \iint_{(\sigma)} \frac{F(N)}{|\cos(n, Z)|} d\sigma_{xy}, \tag{29}$$

We assume here that $\cos(n, Z)$ differs from zero and that the function $F(N)$ has the same value at the point N of (σ) as the given function $F(M)$ has at the point M of (S) whose projection is N. If the equation

of (S) is given explicitly by (22), whilst $F(M)$ is expressed as a function $F(x, y, z)$ of the coordinates, it is sufficient to substitute $z = f(x, y)$ in the expression of $F(x, y, z)$ when integrating over (σ_{xy}), i.e. $F(N) = F[x, y, f(x, y)]$. The denominator on the right-hand side of (29) is given by the third of equations (24).

We note that the properties of double integrals listed in [61], including, in particular, the mean value theorem, apply to an integral over a surface.

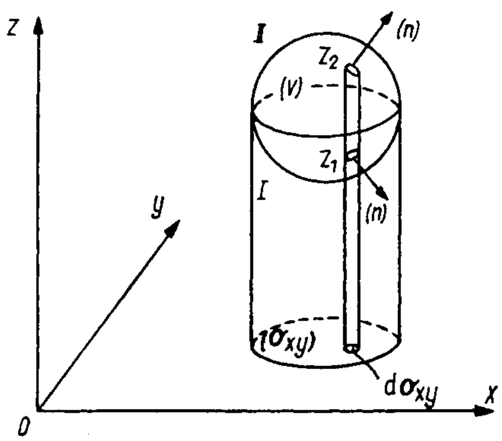

FIG. 54

We now prove Ostrogradskii's formula, which is basic in the theory of multiple integrals, and which establishes the relationship between a triple integral over a volume (V) and the integral over its boundary surface (S). We shall make the same assumption as in [58], that lines parallel to the z axis intersect (S) in not more than two points. We keep the notation of Fig. 44 [58], whilst introducing in addition the direction (n) of the normal to (S), (n) being assumed taken outwards from (V) (the outward normal) (Fig. 54); (n) forms an acute angle with the z axis on the upper part of the surface (II), and an obtuse angle on the lower part (I). Hence, $\cos(n, Z)$ is negative over part (I) and $|\cos(n, Z)| = -\cos(n, Z)$. Formula (26) gives:

$$d\sigma_{xy} = \cos(n, Z)\, dS \text{ on (II)};\quad d\sigma_{xy} = -\cos(n, Z)\, dS \text{ on (I)} \qquad (30)$$

Let $R(x, y, z)$ be continuous in the domain (v) as far as the boundary (S) and let its derivative $\partial R(x, y, z)/\partial z$ be continuous.

We consider the triple integral over (v) of the function $\partial R(x, y, z)/\partial z$. We have, on using (16):

$$\iiint\limits_{(v)} \frac{\partial R(x, y, z)}{\partial z}\, dv = \iint\limits_{(\sigma_{xy})} d\sigma_{xy} \int\limits_{z_1}^{z_2} \frac{\partial R(x, y, z)}{\partial z}\, dz.$$

But the integral of a derivative is equal to the difference in the values of the original function at the upper and lower limits:

$$\iiint\limits_{(v)} \frac{\partial R(x, y, z)}{\partial z}\, dv = \iint\limits_{(\sigma_{xy})} [R(x, y, z_2) - R(x, y, z_1)]\, d\sigma_{xy}$$

or

$$\iiint\limits_{(\sigma)} \frac{\partial R(x, y, z)}{\partial z} \, \mathrm{d}v = \iint\limits_{(\sigma_{xy})} R(x, y, z_2) \, \mathrm{d}\sigma_{xy} - \iint\limits_{(\sigma_{xy})} R(x, y, z_1) \, \mathrm{d}\sigma_{xy}.$$

We reduce integration over (σ_{xy}) to integration over (S) on replacing $\mathrm{d}\sigma_{xy}$ by $\mathrm{d}S$ in accordance with (30); the first of expressions (30) must be used for the first integral, containing the ordinate z_2 of part (II) of (S) as variable, the result being an integral over (II), whilst similarly, the second of expressions (30) is used for the second integral and an integral over (I) obtained:

$$\iiint\limits_{(v)} \frac{\partial R(x, y, z)}{\partial z} \, \mathrm{d}v = \iint\limits_{(\mathrm{II})} R(x, y, z) \cos(n, Z) \, \mathrm{d}S + \\ + \iint\limits_{(\mathrm{I})} R(x, y, z) \cos(n, Z) \, \mathrm{d}S .$$

We can drop the subscript for z since it is implied by the part of the surface over which integration is carried out. The right-hand side contains the sum of integrals over (II) and (I), i.e. amounts to the integral over the whole of (S):

$$\iiint\limits_{(v)} \frac{\partial R(x, y, z)}{\partial z} \, \mathrm{d}v = \iint\limits_{(S)} R(x, y, z) \cos(n, Z) \, \mathrm{d}S . \tag{31}$$

We might have taken another two functions $P(x, y, z)$ and $Q(x, y, z)$ and shown in a similar manner that

$$\iiint\limits_{(v)} \frac{\partial P(x, y, z)}{\partial x} \, \mathrm{d}v = \iint\limits_{(S)} P(x, y, z) \cos(n, X) \, \mathrm{d}S$$

$$\iiint\limits_{(v)} \frac{\partial Q(x, y, z)}{\partial y} \, \mathrm{d}v = \iint\limits_{(S)} Q(x, y, z) \cos(n, Y) \, \mathrm{d}S .$$

We arrive at Ostrogradskii's formula on adding the left- and right-hand sides of the three expressions obtained:

$$\iiint\limits_{(v)} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) \mathrm{d}v = \\ = \iint\limits_{(S)} [P \cos(n, X) + Q \cos(n, Y) + R \cos(n, Z)] \, \mathrm{d}S . \tag{32}$$

The arguments x, y, z of P, Q and R have been omitted for brevity, but it should be borne in mind that these given functions are continuous in the volume (v), together with their derivatives.

A variety of applications of Ostrogradskii's formula will be found in later chapters.

The quantities $\cos(n, X)$, $\cos(n, Y)$, $\cos(n, Z)$ are functions defined on the surface (S) and have been assumed continuous. A more general assumption can be made, that (S) is divisible into a finite number of pieces, on each of which the functions are continuous. This would apply if (S) were, say, a polyhedron.

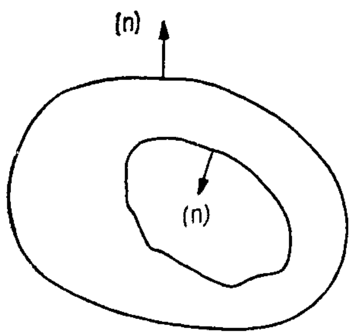

FIG. 55

We supposed when deriving (31) that lines parallel to the z axis cut the surface (S) of (v) in not more than two points. A generalization is easily made for other types of domain. We notice to start with that if (S) includes a lateral cylindrical part with generators parallel to the z axis in addition to the upper part (II) and lower part (I), we have $\cos(n, Z) = 0$ on the lateral part, so that addition of this part to the right-hand side of (31) does not change the value of the integral over the surface; the proof of the formula therefore remains valid throughout. It is sufficient in more general cases to divide (v) into a finite number of sections with the aid of cylindrical surfaces whose generators are parallel to the z axis, such that the previous conditions are satisfied for each section, to which (31) may then be applied. Addition of the results obtained will give us the triple integral over the total volume (v) on the left-hand side; we shall have on the right-hand side the sum of the integrals over the total surface of the sections, whilst the integrals over the cylindrical surfaces that have been added will be zero, as shown above. The result of adding the right-hand sides will, therefore, be the integral over the surface (S) of the original volume (v), and (31) is shown to be valid in general.

We remark that the arguments remain valid when (v) is bounded by several surfaces: by one from the outside and by the remainder from inside. Figure 55 illustrates the case of (v) bounded by two surfaces. The integration on the right-hand side of (31) is now over all the surfaces bounding (v), with the normal (n) directed inwards from the inner surfaces [i.e. outwards from (v)].

64. Integrals over a given side of a surface. Use is occasionally made of a different definition and a different notation for surface integrals. We start with the case when (S), illustrated in Fig. 54, satisfies the conditions laid

down at the beginning of the previous article. Normals can be drawn in two opposite directions at every point of the surface. One direction forms an acute angle, and the other an obtuse angle with the z axis. Correspondingly, we can distinguish between *an upper and a lower side* of the surface. Let $R(x, y, z)$ be given on (S), as above. We take the integral:

$$\iint_{(S)} R \cos(n, Z)\, dS. \tag{33}$$

The value of the integral depends on the choice of the direction of the normal or what amounts to the same thing, on specifying the side of the surface over which the integration is carried out. On integrating over the upper side, $\cos(n, Z) > 0$, and $\cos(n, Z)\, dS = d\sigma_{xy}$, whilst over the lower side, $\cos(n, Z) <$ < 0, and $\cos(n, Z)\, dS = -d\sigma_{xy}$, where $d\sigma_{xy}$ is the projection of an element of (S) on the XY plane, i.e. an element of the area (σ) in (29). We can write $d\sigma_{xy} = dx\, dy$ in (x, y) coordinates, so that integral (33) reduces to an integral over (σ) in the XY plane:

$$\iint_{(\sigma)} R[x, y, f(x, y)]\, dx\, dy \quad \text{or} \quad -\iint_{(\sigma)} R[x, y, f(x, y)]\, dx\, dy, \tag{34}$$

depending on which side of the surface the integration is made. These are in fact often written in the same way, as

$$\iint_{(S)} R\, dx\, dy, \tag{35}$$

whilst indicating which side of the surface is concerned. For instance, if integration is over the lower side of (S), (35) amounts to the second of integrals (34). We can define (35) directly as the limit of the sum $\Sigma R(M_k)\, \Delta\sigma_k$, where $R(M_k)$ is the value of R at the point M_k of the element ΔS_k of (S) whose corresponding projection on the XY plane is $\Delta\sigma_k$, the $\Delta\sigma_k$ being reckoned positive if integration is over the upper side of the surface and negative if over the lower side.

We now take the general case of (S), and let M_0 be a given point of the surface. We specify a definite direction for the normal (n) at this point; starting from M_0 and moving continuously over (S), we examine the continuous variation in the direction of (n). If any continuous movement leads to a definite direction of the normal at any point of the surface, it is said to be two-sided. If we specified differently the direction of (n) at the initial point M_0, continuous movement would lead us to an opposite direction of (n) at every other point of the surface. This allows us to speak of the two sides of (S), depending on the direction specified for (n) at M_0 and therefore specified at the remaining points. Having fixed the side of the surface, we get a definite value for integral (33), which may be written in the form (35) on indicating the side concerned.

We can similarly define the integrals:

$$\iint_{(S)} P\, dy\, dz \quad \text{and} \quad \iint_{(S)} Q\, dx\, dz.$$

where $P(x, y, z)$ and $Q(x, y, z)$ are functions given on (S). These integrals coincide respectively with:

$$\iint_{(S)} P \cos(n, X)\, dS \quad \text{and} \quad \iint_{(S)} Q \cos(n, Y)\, dS.$$

With these definitions, (32) can now be written as

$$\iiint\limits_{(v)} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) \mathrm{d}v = \iint\limits_{(S)} (P \,\mathrm{d}y \,\mathrm{d}z + Q \,\mathrm{d}x \,\mathrm{d}z + R \,\mathrm{d}x \,\mathrm{d}y)$$

where the integration on the right-hand side is over the outer side of (S).

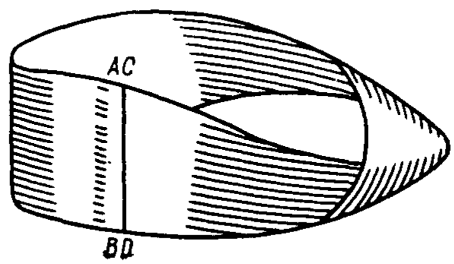

FIG. 56

We note the existence of one-sided surfaces, where continuous movement over the surface implies continuous variation in the direction of the normal such that, on returning to the initial point, the new direction of the normal can be the opposite of the original direction. A simple example is the Möbius strip, obtained by taking a rectangular sheet of paper $ABCD$, twisting it once and fastening edge AB to edge CD in such a way that A coincides with C and B with D (Fig. 56). Both sides of the ring thus obtained could be painted, without removing the brush from the surface, i.e. without passing over its boundary.

65. Moments. An application of multiple integrals is to the theory of the moments of different orders of mechanical systems. Let a system of n material particles of masses $m_1, m_2, \ldots, m_n$ be given at points:

$$M_1, M_2, \ldots, M_n,$$

respectively.

The k-th order moment of the system with respect to a plane (Δ), *a straight line* (d) *or a point* (D) *is defined as the sum of the products of the mass of each particle of the system with the k-th power of the distance from* (Δ), (d) *or* (D):

$$\sum_{i=1}^{n} r_i^k m_i.$$

The zero order moment becomes from this view-point simply the total mass of the system:

$$m = \sum_{i=1}^{n} m_i.$$

The first order moment with respect to the plane (Δ) is called the *statical moment of the system with respect to the plane*. We encounter the statical moments relative to the coordinate planes in the expressions for *the coordinates of the centre of gravity of the system:*

$$x_g = \frac{\sum\limits_{i=1}^{n} m_i x_i}{m}; \quad y_g = \frac{\sum\limits_{i=1}^{n} m_i y_i}{m}; \quad z_g = \frac{\sum\limits_{i=1}^{n} m_i z_i}{m}. \tag{36}$$

The distances x_i, y_i, z_i from the coordinate planes are reckoned algebraically here, i.e. as positive or negative.

The *second order moments* are usually called *moments of inertia* of the system. The expressions

$$\sum_{i=1}^{n} x_i^2 m_i; \quad \sum_{i=1}^{n} y_i^2 m_i; \quad \sum_{i=1}^{n} z_i^2 m_i$$

represent the moments of inertia of the system with respect to the coordinate planes; the expressions

$$\sum_{i=1}^{n} (y_i^2 + z_i^2)\, m_i; \quad \sum_{i=1}^{n} (z_i^2 + x_i^2)\, m_i; \quad \sum_{i=1}^{n} (x_i^2 + y_i^2) m_i$$

represent the moments of inertia relative to the x, y, z axes; and finally,

$$\sum_{i=1}^{n} (x_i^2 + y_i^2 + z_i^2)\, m_i$$

is the moment of inertia relative to the origin O.

The expressions

$$\sum_{i=1}^{n} y_i\, z_i\, m_i; \quad \sum_{i=1}^{n} z_i\, x_i\, m_i; \quad \sum_{i=1}^{n} x_i\, y_i\, m_i,$$

are occasionally encountered in addition to the above, and are referred to as the *products of inertia of the system relative to the x, y, z axes.*

If we are dealing with continuous distributions of matter instead of systems of a finite number of particles, the above sums have to be replaced by definite integrals; these are single, double or triple, according as the distribution of mass is linear, over a surface or over a volume; instead of the m_i, we now have the density $f(M)$ at a given point M multiplied by the element of line, surface or volume.

For example, the moment of inertia of a three-dimensional domain (v) relative to axis OX is given by the triple integral

$$\iiint_{(v)} (y^2 + z^2)\, f(M)\, \mathrm{d}v.$$

If $f(M)$ is the constant density f_0, this can be taken outside the integration sign; an f_0 could be cancelled out in (36), leaving functions of x, y and z under the integrals in the numerators, and the volume or area of the total domain in the denominator.

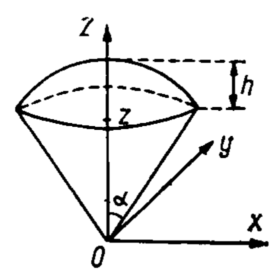

FIG. 57

Examples. **1.** The centre of gravity of a sector of a homogeneous sphere (Fig. 57). On using the coordinate system shown in the figure, we need only find the ordinate

$$z_g = \frac{\iiint_{(v)} z\, \mathrm{d}v}{v}.$$

We have here:

$$v = \int_0^{2\pi} d\varphi \int_0^{\alpha} \sin\theta \, d\theta \int_0^{a} \varrho^2 \, d\varrho = \frac{2}{3}\pi a^3 (1 - \cos\alpha) = \frac{2}{3}\pi a^2 h$$

$$\iiint_{(v)} z \, dv = \int_0^{2\pi} d\varphi \int_0^{\alpha} \sin\theta \, d\theta \int_0^{a} \varrho \cos\theta \varrho^2 \, d\varrho = 2\pi \int_0^{\alpha} \sin\theta \cos\theta \, d\theta \int_0^{a} \varrho^3 \, d\varrho =$$

$$= \frac{\pi}{8} a^4 (1 - \cos 2\alpha)$$

$$z_g = \frac{3}{16} \; a \; \frac{1 - \cos 2\alpha}{1 - \cos\alpha} = \frac{3}{8} a(1 + \cos\alpha) = \frac{3}{8}(2a - h),$$

where a is the radius of the sphere.

2. If the mass is assumed to be distributed only on the spherical surface (S) of the sector, the ordinate of the centre of gravity becomes

$$z_g = \frac{\int_{(S)} z \, ds}{s},$$

where s is the area of surface (S). The equation of the surface is here $x^2 + y^2 + z^2 = a^2$ or $z = \sqrt{a^2 - (x^2 + y^2)}$, and it is easily shown that

$$\cos(n, z) = \frac{1}{\sqrt{1 + p^2 + q^2}} = \frac{z}{a},$$

so that

$$\iint_{(S)} z \, ds = \iint_{(\sigma_{xy})} z \; \frac{d\sigma_{xy}}{\cos(n, z)} = a \iint_{(\sigma_{xy})} d\sigma_{xy} = \pi a^3 \sin^2\alpha,$$

where (σ_{xy}) is clearly a circle with centre at the origin and radius $a \sin\alpha$.

The area s will be:

$$s = \iint_{(\sigma_{xy})} \sqrt{1 + p^2 + q^2} \, d\sigma_{xy} = a \iint_{(\sigma_{xy})} \frac{d\sigma_{xy}}{\sqrt{a^2 - (x^2 + y^2)}} =$$

$$= a \int_0^{2\pi} d\varphi \int_0^{a \sin\alpha} \frac{r \, dr}{\sqrt{a^2 - r^2}} = 2\pi a^2 (1 - \cos\alpha),$$

and finally:

$$z_g = \frac{\pi a^3 \sin^2\alpha}{2\pi a^2 (1 - \cos\alpha)} = a \cos^2 \frac{\alpha}{2}.$$

We had in the previous example the smaller value for z_g:

$$\frac{3}{8} a(1 + \cos\alpha) = \frac{3}{4} a \cos^2 \frac{\alpha}{2}.$$

3. If the centre of gravity is at the origin, all the statical moments are zero, as follows at once from the relationships:

$$\iiint\limits_{(v)} xf\,\mathrm{d}v = mx_g;$$

$$\iiint\limits_{(v)} yf\,\mathrm{d}v = my_g;$$

$$\iiint\limits_{(v)} zf\,\mathrm{d}v = mz_g.$$

4. The moments of inertia of a homogeneous right circular cylinder (Fig. 58) relative to the axis of the cylinder and to the diameter of its central section. Taking f_0 as the constant density, we have:

$$J_z = f_0 \iiint\limits_{(v)} r^2 r\,\mathrm{d}r\,\mathrm{d}\varphi\;\mathrm{d}z = 2f_0 \int\limits_0^{2\pi} \mathrm{d}\varphi \int\limits_0^{a} r^3\,\mathrm{d}r \int\limits_0^{h} \mathrm{d}z = \pi a^4 h f_0 = m\frac{a^2}{2};$$

$$J_x = f_0 \iiint\limits_{(v)} (z^2 + r^2\sin^2\varphi)\, r\,\mathrm{d}r\,\mathrm{d}\varphi\,\mathrm{d}z = 2f_0 \int\limits_0^{2\pi} \mathrm{d}\varphi \int\limits_0^{h} \mathrm{d}z \int\limits_0^{a} (z^2 + r^2\sin^2\varphi)\, r\,\mathrm{d}r =$$

$$= 2f_0 \int\limits_0^{2\pi} \mathrm{d}\varphi \int\limits_0^{h} z^2\,\mathrm{d}z \int\limits_0^{a} r\,\mathrm{d}r + 2f_0 \int\limits_0^{2\pi} \sin^2\varphi\,\mathrm{d}\varphi \int\limits_0^{h} \mathrm{d}z \int\limits_0^{a} r^2\,\mathrm{d}r = \frac{2}{3}\pi h^3 a^2 f_0 +$$

$$+ \frac{\pi}{2} h a^4 f_0 = m\left(\frac{h^2}{3} + \frac{a^2}{4}\right),$$

where $2h$ is the height of the cylinder, a is the radius of its base and m its mass

5. The moments of inertia of the homogeneous ellipsoid:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

We let f_0 denote the density and have, on dividing into layers parallel to the XOY plane:

$$J_{xy} = f_0 \iiint\limits_{(v)} z^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = f_0 \int\limits_{-c}^{+c} z^2 \pi ab\left(1 - \frac{z^2}{c^2}\right)\mathrm{d}z =$$

$$= 2\pi ab\, f_0\left(\frac{c^3}{3} - \frac{c^3}{5}\right) = m\frac{1}{5}c^2.$$

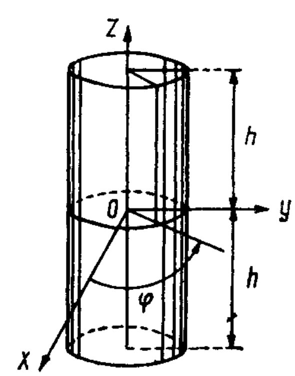

FIG. 58

We can change the symbols cyclically to find that

$$J_{yz} = m \cdot \frac{1}{5} a^2; \quad J_{zx} = m \cdot \frac{1}{5} b^2$$

$$J_x = J_{xy} + J_{xz} = m \cdot \frac{1}{5} (b^2 + c^2)$$

$$J_y = m \cdot \frac{1}{5} (c^2 + a^2); \quad J_z = m \cdot \frac{1}{5} (a^2 + b^2)$$

$$J_0 = J_{xy} + J_{yz} + J_{zx} = m \cdot \frac{1}{5} (a^2 + b^2 + c^2).$$

6. The kinetic energy of rotation of a rigid body about an axis (δ).

We know that the velocity V of any point of the body is equal to the angular velocity ω of the body multiplied by the distance of the point from the axis of rotation (δ). We find the kinetic energy of the body by dividing it into elements of mass Δm and finding the kinetic energy ΔT of each; then

$$T = \sum \Delta T.$$

Since the element Δm is small, its mass can be considered to be concentrated at any point M of the element; the kinetic energy ΔT of Δm is then

$$\Delta T = \frac{1}{2} V^2 \Delta m = \frac{1}{2} \omega^2 r_\delta^2 f(M)\, \Delta v,$$

where $f(M)$ is the density of the body at M and r_δ is the distance of M from the axis (δ). Hence we obtain, by definition of a triple integral:

$$T = \iiint\limits_{(v)} \frac{1}{2} \omega^2 r_\delta^2 f(M)\, \mathrm{d}v = \frac{1}{2} \omega^2 J_\delta,$$

where

$$J_\delta = \iiint\limits_{(v)} r_\delta^2 f(M)\, \mathrm{d}v$$

is the moment of inertia of the body about the axis of rotation (δ).

Remark. It is sometimes more convenient to evaluate the volume or the moments of a body in terms of double or even single instead of triple integrals. This comes from the fact that, when a triple integral is represented as the single integral of a double integral, or as the double integral of a single integral, the interior integral may occasionally be evaluated directly from elementary considerations without performing an actual integration. This creates the impression that a double or single, instead of a triple integral is used for the calculation.

We can take as an example the moment of inertia J_{xy} of a body (v) relative to the XY plane, where (v) is bounded by the planes $z = 0$, $z = h$ and by the surface formed by revolution of the curve $x = f(z)$ about the z axis. The moment can be obtained from a single integral if the body is imagined to consist of plane circular discs, parallel to the XY plane; the volume of such a disc is $\pi[f(z)]^2\, \mathrm{d}z$, and we can write:

$$J_{xy} = \pi \int\limits_0^h z^2 [f(z)]^2\, \mathrm{d}z.$$

The same moment of inertia is given by the triple integral:

$$J_{xy} = \iiint\limits_{(v)} z^2\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int\limits_0^h z^2\,\mathrm{d}z \iint\limits_{(\sigma_z)} \mathrm{d}x\,\mathrm{d}y,$$

where (σ_z) is the section cut from (v) by a plane parallel to the XY plane and distant z from it. The interior double integral gives the area (σ_z), i.e. is equal to $\pi[f(z)]^2$.

§ 7. Line integrals

66. Definition of a line integral. Let (l) be a curve with a definite direction in space (Fig. 59), A being the initial point and B

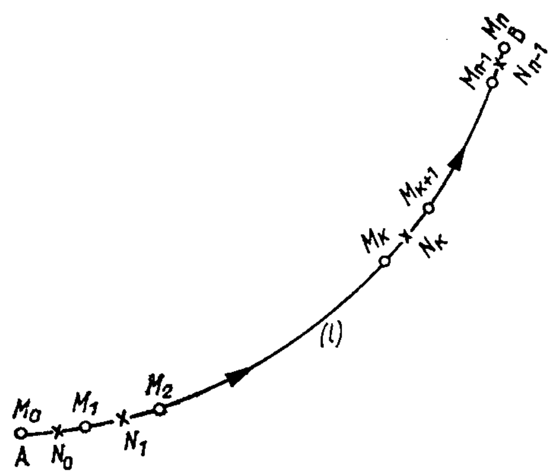

FIG. 59

the end-point of the curve. We shall measure the length of arc along (l) from the initial point A. Let a continuous function $f(M)$ be given on (l), and let (l) be divided into n parts by the points $M_0, M_1, \ldots, M_{n-1}, M_n$, where M_0 coincides with A and M_n with B. We take any point N_k of each segment $M_k M_{k+1}$ $(k = 0, 1, \ldots, n-1)$ and form the sum $\sum\limits_{k=0}^{n-1} f(N_k)\,\Delta s_k$, where Δs_k is the length of arc $M_k M_{k+1}$ of (l). The limit of this sum on indefinite increase in the number n of divisions and indefinite decrease of each of the $M_k M_{k+1}$ is called the line integral of the function $f(M)$ over (l) and is written:

$$\int\limits_{(l)} f(M)\,\mathrm{d}s = \lim \sum_{k=1}^{n-1} f(N_k)\,\Delta s_k. \tag{1}$$

The position of the variable point M on (l) is fully defined by the length of arc $s = \smile AM$, and $f(M)$ can, therefore, be taken as a function of the independent variable s, i.e. $f(M) = f(s)$; integral (1) is an ordinary definite integral:

$$\int_{(l)} f(M)\,\mathrm{d}s = \int_0^l f(s)\,\mathrm{d}s,$$

where l is the length of arc of (l). We note that (l) can be a closed curve, i.e. B can coincide with A.

So far, we have not used the fact that (l) has a specified direction, which becomes important later. Let the coordinates of M be (x, y, z), referred to a system of rectangular axes in space. Let $P(x, y, z)$ be a continuous function along (l), and let (ξ_k, η_k, ζ_k) denote the coordinates of N_k, and Δx_k the projection of the directional arc $\overline{M_k M_{k+1}}$ on the x axis; Δx_k can of course be positive, negative, or zero. We now form the sum of the products of $P(N_k) = P(\xi_k, \eta_k, \zeta_k)$ with Δx_k instead of Δs_k, i.e. the sum

$$\sum_{k=0}^{n-1} P(\xi_k, \eta_k, \zeta_k)\,\Delta x_k.$$

The limit of this sum is called the line integral of $P(x, y, z)$ over (l) and is written

$$\int_{(l)} P(x, y, z)\,\mathrm{d}x = \lim \sum_{k=0}^{n-1} P(\xi_k, \eta_k, \zeta_k)\,\Delta x_k.$$

The integrals

$$\int_{(l)} Q(x,y,z)\,\mathrm{d}y \quad \text{and} \quad \int_{(l)} R(x, y, z)\,\mathrm{d}z$$

are similarly defined, where $Q(x, y, z)$ and $R(x, y, z)$ are continuous functions along (l). Addition of these three integrals gives us the general form of the line integral, which is written:

$$\int_{(l)} P(x, y, z)\,\mathrm{d}x + Q(x, y, z)\,\mathrm{d}y + R(x, y, z)\,\mathrm{d}z. \tag{2}$$

Integral (2) is by definition the limit of a sum of the form:

$$\sum_{k=0}^{n-1} \left[P(\xi_k, \eta_k, \zeta_k)\,\Delta x_k + Q(\xi_k, \eta_k, \zeta_k)\,\Delta y_k, + R(\xi_k, \eta_k, \zeta_k)\,\Delta z_k\right], \tag{3}$$

where Δy_k, Δz_k are the projections of $\overline{M_k M_{k+1}}$ on the y and z axes respectively. The connection between integrals of type (2) and type (1)

is easily found. The coordinates (x, y, z) of the variable point M of curve (l) can be taken as functions of the length of arc $s = \smile AM$. As we know from [I, 160], the derivatives of these functions are the direction-cosines of the tangent to (l), i.e.

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \cos(t, X); \qquad \frac{\mathrm{d}y}{\mathrm{d}s} = \cos(t, Y); \qquad \frac{\mathrm{d}z}{\mathrm{d}s} = \cos(t, Z),$$

where t is the tangent to (l) at M, directed in the same sense as (l). The symbol of the type (α, β) denotes as usual the angle between the directions α and β; the cosine of the angle is independent of the sense in which it is measured, and the sense is not specified here. To an accuracy of higher order infinitesimals we can take:

$$\Delta x_k = \cos(t_k, X)\,\Delta s_k \quad \Delta y_k = \cos(t_k, Y)\,\Delta s_k, \quad \Delta z_k = \cos(t_k, Z)\,\Delta s_k,$$

where t_k is the direction of the tangent at N_k, and integral (2) reduces as the limit of the sum (3) to form (1):

$$\int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z =$$

$$= \int_{(l)} \left[P\cos(t, X) + Q\cos(t, Y) + R\cos(t, Z)\right] \mathrm{d}s, \tag{4}$$

where P, Q, R can be considered as functions of s along (l).

Let the equation of (l) be given in parametric form:

$$x = \varphi(\tau); \quad y = \psi(\tau); \quad z = \omega(\tau), \tag{5}$$

where the point (x, y, z) describes the curve (l) from A to B as τ varies from a to b. We shall suppose functions (5) continuous, with continuous first order derivatives, in the closed interval (a, b), whilst taking $a < b$ for clarity.

Let $\tau = \tau_k$ correspond to M_k, and let us consider the first of the summations (3). Let $\tau = \tau_k'$ correspond to the point (ξ_k, η_k, ζ_k) of the curve. We can write, using the formula of finite increments [I, 63]:

$$\Delta x_k = \varphi(\tau_{k+1}) - \varphi(\tau_k) = \varphi'(\tau_k'')(\tau_{k+1} - \tau_k),$$

where τ_k'' is a value of τ in the interval (τ_k, τ_{k+1}). Our sum can now be written as:

$$\sum_{k=0}^{n-1} P(\xi_k, \eta_k, \zeta_k)\,\Delta x_k =$$

$$= \sum_{k=1}^{n-1} P\left[\varphi(\tau_k'),\ \psi(\tau_k'),\ \omega(\tau_k')\right] \varphi'(\tau_k'')(\tau_{k+1} - \tau_k). \tag{6}$$

This closely resembles the sum:

$$\sigma = \sum_{k=0}^{n-1} P\left[\varphi(\tau_k''), \psi(\tau_k''), \omega(\tau_k'')\right] \varphi'(\tau_k'')(\tau_{k+1} - \tau_k),$$

which tends in the limit, as the greatest of the $(\tau_{k+1} - \tau_k)$ tends to zero, to the definite integral:

$$\int_a^b P\left[\varphi(\tau), \psi(\tau), \omega(\tau)\right] \varphi'(\tau)\, \mathrm{d}\tau. \tag{7}$$

We now show that the difference between σ and (6) tends to zero. It will follow immediately from this that (6) has a limit, equal to integral (7). The difference is of the form:

$$\eta = \sum_{k=0}^{n-1} \left\{P[\varphi(\tau_k'), \psi(\tau_k'), \omega(\tau_k')] - \right.$$
$$\left. - P\left[\varphi(\tau_k''), \psi(\tau_k''), \omega(\tau_k'')\right]\right\} \varphi'(\tau_k'')(\tau_{k+1} - \tau_k).$$

Since τ_k' and τ_k'' belong to (τ_k, τ_{k+1}) and $P[\varphi(\tau), \psi(\tau), \omega(\tau)]$ is uniformly continuous, for any small positive ε there exists a δ such that [I, 32]

$$\left| P\left[\varphi(\tau_k'), \psi(\tau_k'), \omega(\tau_k')\right] - P\left[\varphi(\tau_k''), \psi(\tau_k''), \omega(\tau_k'')\right] \right| < \varepsilon,$$

provided $(\tau_{k+1} - \tau_k) < \delta$. The absolute value of η will then satisfy

$$|\eta| < \varepsilon \sum_{k=0}^{n-1} |\varphi'(\tau_k'')| (\tau_{k+1} - \tau_k).$$

The function $\varphi'(\tau)$ is continuous and therefore bounded in (a, b), i.e. $|\varphi'(\tau)| < K$, where K is a definite number [I, 35]. Hence we have:

$$|\eta| < \varepsilon K \sum_{k=0}^{n-1} (\tau_{k+1} - \tau_k) = \varepsilon K (b - a).$$

Since $\varepsilon \to 0$ if $\max(\tau_{k+1} - \tau_k) \to 0$, it can be seen that η in fact tends to zero, and (7) is the limit of summation (6). The remainder of summations (3) can be considered in the same way and hence we can show that, with the assumptions made, integral (2) can be written as an ordinary definite integral:

$$\int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z = \int_a^b \left[P\varphi'(\tau) + Q\psi'(\tau) + R\omega'(\tau)\right] \mathrm{d}\tau, \tag{8}$$

where P, Q and R must be expressed in terms of τ in accordance with (5).

Several of the properties of simple integrals listed in [I, 94] may be generalized at once for line integrals. For instance:

I. If the curve (l) consists of distinct portions (l_1), (l_2), ..., (l_m),

$$\int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z = \int_{(l_1)} [P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z] + \\ + \int_{(l_2)} [P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z] + \ldots + \int_{(l_m)} [P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z].$$

II. The value of a line integral depends on the direction of the curve (l), as well as on the form of (l) and the integrand; however, *when the direction of the curve of integration is changed, the integral only changes sign.*

If (l) as a whole does not satisfy the conditions stated above, but can be divided into a finite number of sections, each of which has parametric equations (5), we can apply (7) to each section, and the integral over the whole curve can be written as the sum of the integrals over the separate sections. It may easily be shown that this is equivalent to the limit of summations (3) for the whole curve. We shall only consider in future curves (l) which satisfy the conditions just mentioned. Finally, we notice that if τ is the length of arc $s = \smile AM$, formula (8) reduces to (4).

If (l) is a plane curve, say in the XY plane, (2) becomes:

$$\int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y,$$

where P and Q are functions of (x, y) given along (l).

67. Work done by a field of force. Examples. Calculations of work done lead naturally to line integrals of type (2). Let a point M describe a trajectory (l) under the action of a force $\mathbf{F}$, which is a function of points along (l). The work done is found by dividing (l) into small segments and considering one of these, say M_k, M_{k+1}. In view of the smallness of the segment, $\mathbf{F}$ can be reckoned to maintain constant over the segment its value at M_k, whilst arc $\smile M_k M_{k+1}$ can be replaced by chord $M_k M_{k+1}$. The work done over the segment is thus given approximately by

$$\Delta E_k \sim |\mathbf{F}_k|\,|\overline{M_k M}_{k+1}| \cos(\mathbf{F}_k, \overline{M_k M}_{k+1}),$$

where $|\mathbf{F}_k|$ denotes the length of vector $\mathbf{F}$ at M_k, $|\overline{M_k M}_{k+1}|$ is the length of $\overline{M_k M}_{k+1}$; and ΔE_k is the work done over the segment

$\smile M_k M_{k+1}$. On using the formula of analytical geometry for the angle between two directions, we can write:

$$\Delta E_k \sim |\mathbf{F}_k| \, |\overline{M_k M}_{k+1}| \left[\cos(\mathbf{F}_k, X)\cos(\overline{M_k M}_{k+1}, X) + \right.$$
$$\left. + \cos(\mathbf{F}_k, Y)\cos(\overline{M_k M}_{k+1}, Y)\cos(\mathbf{F}_k, Z)\ \cos(\overline{M_k M}_{k+1}, Z)\right],$$

or, on removing the brackets and denoting the projections of $\mathbf{F}$ on the coordinate axes P, Q, R:

$$\Delta E_k \sim P_k \Delta x_k + Q_k \Delta y_k + R_k \Delta z_k,$$

where the subscripts of P, Q, R indicate that their values are taken at M_k. We now sum over all the segments and pass to the limit, so that the work done is accurately:

$$E = \int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z.$$

Examples. **1.** The work done by the constant force of gravity when a point M of mass m is displaced via any curve (l) from the position $M_1(a_1, b_1, c_1)$ to $M_2(a_2, b_2, c_2)$ is given by the integral

$$\int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z = \int_{c_1}^{c_2} mg\,\mathrm{d}z = mg\,(c_2 - c_1)$$

(the z axis is directed vertically downwards); hence it follows that the work depends only on the initial and final positions of the point, and not on the path traversed. Here we have an example of a line integral the value of which depends only on the initial and final points of integration, and not on the path.

2. The work done by the gravitational force when a unit mass is displaced from the point M_1 to M_2 in its attraction to a fixed centre of mass m. If we take the centre at the origin and let r be the radius vector to the point, F can be seen to be in the opposite direction to $\overline{OM}$ and equal to fm/r^2, where f is the gravitational constant. Hence we have here:

$$P = -\frac{fm}{r^2}\cdot\frac{x}{r}; \qquad Q = -\frac{fm}{r^2}\cdot\frac{y}{r}; \qquad R = -\frac{fm}{r^2}\cdot\frac{z}{r}$$

$$E = -fm\int_{(l)} \frac{x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z}{r^3} = -fm\int_{(l)} \frac{r\,\mathrm{d}r}{r^3} = fm\int_{(l)} \mathrm{d}\left(\frac{1}{r}\right),$$

and if r_2, r_1 denote the respective distances of M_2 and M_1 from the gravitational centre,

$$E = fm\left(\frac{1}{r_2} - \frac{1}{r_1}\right),$$

and here the work, i.e. the corresponding line integral, depends only on the initial and final points, and not on the path.

If we introduce the potential of the point mass,

$$U = \frac{fm}{r},$$

so that

$$P = \frac{\partial U}{\partial y}; \qquad Q = \frac{\partial U}{\partial y}; \qquad R = \frac{\partial U}{\partial z},$$

the work E is given by the difference in the values of U at M_2 and M_1, i.e.

$$E = U(M_2) - U(M_1).$$

We consider line integrals over plane curves in the next examples.

3. We consider the established plane flow of an incompressible fluid of constant density which we take as unity. With this, the velocity of motion $\mathbf{v}$

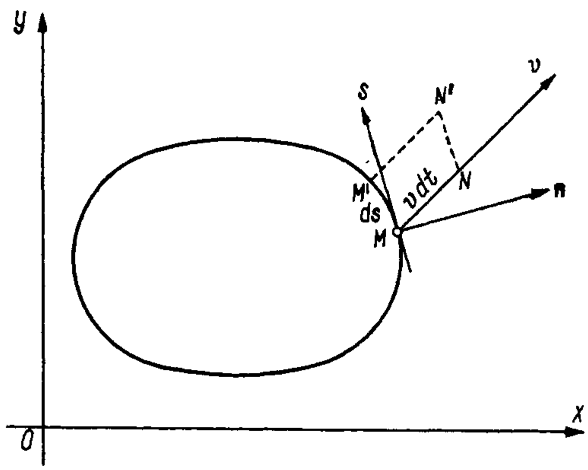

FIG. 60

of a fluid particle situated at $M(x, y)$ depends only on (x, y). We find the quantity of fluid q flowing in unit time past a given contour (l) (Fig. 60). Let u and v denote the projections of $\mathbf{v}$ on the coordinate axes. We take an element $\smile MM' = ds$ of (l), and assume that the velocities of all the particles of the element are approximately equal; it can then be seen that, during a small interval of time dt, all the particles move an amount $|\mathbf{v}|\,dt$ in the direction of $\mathbf{v}$ to the position NN'. The area of the parallelogram $MNN'M'$ can be written as the product of the base ds and the projection of $\mathbf{v}\,dt$ on the outward normal (n) to (l), i.e.

$$\text{area } MNN'M' = |\mathbf{v}| \cos(v, n)\, dt\, ds,$$

where $|\mathbf{v}|$ is the length of $\mathbf{v}$. If (s) is the direction of the tangent to (l) on a counter-clockwise circuit, we have:

$$(n, X) = (s, Y); \qquad (n, Y) = (s, X) - \pi, \tag{9}$$

where a symbol of the type (α, β) denotes the angle measured counter-clockwise from direction α to β. Hence:

$$\cos(n, X) = \cos(s, Y) \text{ and } \cos(n, Y) = -\cos(s, X).$$

But we know that the angle between two directions is given by

$$\cos(\mathbf{v}, n) = \cos(\mathbf{v}, X)\cos(n, X) + \cos(\mathbf{v}, Y)\cos(n, Y),$$

so that by (9):

$$\cos(\mathbf{v}, n) = \cos(\mathbf{v}, X)\cos(s, Y) - \cos(\mathbf{v}, Y)\cos(s, X).$$

On substituting in the expression for the area and noting that

$$|\mathbf{v}|\cos(v, X) = u; \qquad |\mathbf{v}|\cos(v, Y) = v$$

$$\mathrm{d}s\cos(s, X) = \Delta x; \qquad \mathrm{d}s\cos(s, Y) = \Delta y,$$

we get finally:

$$\text{area } MNN'M' = (-v\,\Delta x + u\,\Delta y)\,\mathrm{d}t.$$

If (v, n) is an obtuse angle here, $\cos(v, n)$ will be negative and an area will be obtained with a negative sign, which corresponds to the case when the fluid *flows into* the domain bounded by (l).

The total quantity of fluid flowing past (l) in time $\mathrm{d}t$ becomes

$$\mathrm{d}t \sum (-v\,\Delta x + u\,\Delta y) = \mathrm{d}t \int_{(l)} -v\,\mathrm{d}x + u\,\mathrm{d}y,$$

and the quantity in unit time is

$$q = \int_{(l)} [-v\,\mathrm{d}x + u\,\mathrm{d}y], \tag{10}$$

with the circuit round (l) taken counter-clockwise. We notice that (l) can be closed. The quantity q given by (10) has the $(+)$ sign if the flow is in the direction of the normal (n), and the $(-)$ sign if in the opposite direction.

We remarked above on the direction of (n); it is bound up with the direction of the integration round (l) and with the orientation of the x and y axes in accordance with (9). If (l) is a closed contour and integration is carried out counter-clockwise (Fig. 60), q gives the difference in the quantities of fluid flowing in and out of the domain bounded by (l) in unit time, where either term of the difference can be missing.

If (l) contains neither sources (which produce fluid: *positive sources*) nor sinks (which absorb fluid: *negative sources*), q must be zero, since otherwise the quantity of fluid inside (l) would increase or diminish, contradicting the properties of incompressibility and absence of sources.

It follows that *the established plane flow of an incompressible fluid is characterized by the equation*

$$\int_{(l)} [-v\,\mathrm{d}x + u\,\mathrm{d}y] = 0. \tag{11}$$

which must be satisfied for every closed contour (l) *not containing sources.*

4. The state of a body in thermodynamics is given by three physical magnitudes: the pressure p, volume v and temperature T (absolute). These are connected by a relationship of the type

$$f(v, p, T) = 0;$$

in the case of an ideal gas, for instance, we have Clapeyron's equation:

$$pv - RT = 0.$$

The state of a body is thus defined by two of the three magnitudes, say by p and v, i.e. by the point $M(p, v)$ on the pv plane.

If the state changes, the point M must describe a curve on the pv plane, which is called a diagram of the process concerned; if the body returns to its original state, we speak of a cyclical process or cycle, and its diagram is a closed curve (l).

We find the quantity of heat Q absorbed by a body during the process by dividing this up into small elementary processes in which small changes Δp, Δv, ΔT occur in p, v, and T. If only one of these magnitudes changes, the quantity of heat absorbed is roughly proportional to the increment of the corresponding variable, whilst if all three change together, the total increment ΔQ is equal to the sum of the component increments, by the principle of addition of small operations [I, 68]. In other words, we have an approximate equation of the form

$$\Delta Q \sim A\, \Delta p + B\, \Delta v + C\, \Delta T$$

and we get finally:

$$Q = \sum \Delta Q = \int A\, \mathrm{d}p + B\, \mathrm{d}v + C\, \mathrm{d}T. \tag{12}$$

On using the equation of state to express T in terms of v and p, we get:

$$T = \varphi(v, p); \qquad \mathrm{d}T = \frac{\partial \varphi}{\partial p}\, \mathrm{d}p + \frac{\partial \varphi}{\partial v}\, \mathrm{d}v;$$

and finally, on substituting these values for T and $\mathrm{d}T$ in the right-hand side of (12), we find:

$$Q = \int_{(l)} P\, \mathrm{d}p + V\, \mathrm{d}v,$$

where P and V are known functions of v and p.

5. Let us consider the expansion or compression of the gas (steam) in the cylinder of a gas (steam) engine. The change in volume Δv is proportional to the displacement of the piston in the cylinder due to the pressure p, so that the work ΔE done by the pressure in producing the change in volume is given by $p\Delta v$ with suitable choice of units; hence the total work done during one cycle is

$$E = \int_{(l)} p\, \mathrm{d}v.$$

68. Areas and line integrals. We find the area σ of the domain (σ), situated in the XY plane and bounded by the closed curve (l). We assume for simplicity (Fig. 61) that a line parallel to

the axis OY cuts (l) in not more than two points. Let y_1 and y_2 be respectively the ordinates of the points of entry into and exit from (σ) of a line parallel to OY, and let a and b be the extreme values of the abscissae of points of (l). We have [I, 101]:

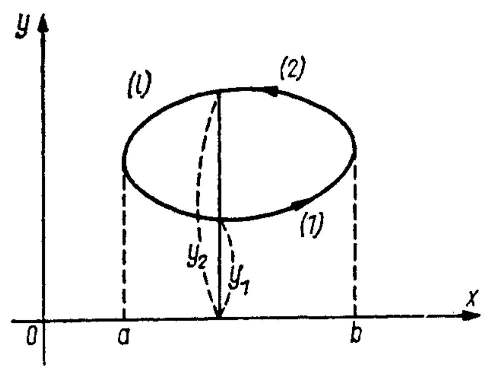

FIG. 61

$$\sigma = \int_a^b (y_2 - y_1)\, dx.$$

Let the so-called points of entry lie on section (1) of the curve, and similarly, those of exit on section (2). The integral:

$$\int_a^b y_2\, dx$$

is identical with the line integral

$$\int_{(2)} y\, dx,$$

with direction from the point $x = a$ to $x = b$, and taken with the reverse sign. Similarly,

$$\int_a^b y_1\, dx,$$

is identical with

$$\int_{(1)} y\, dx,$$

taken from $x = a$ to $x = b$.

Finally we have:

$$\sigma = \int_a^b y_2\, dx - \int_a^b y_1\, dx = -\left[\int_{(1)a}^{b} y\, dx + \int_{(2)b}^{a} y\, dx\right] = -\int_{(l)} y\, dx, \quad (13)$$

where (l) runs in a counter-clockwise direction.

We find by a similar method that

$$\sigma = \int_{(l)} x\, dy. \quad (14)$$

We also get, on adding and dividing by two:

$$\sigma = \frac{1}{2}\int_{(l)} x\, dy - y\, dx. \quad (15)$$

We deduced (13) by assuming that a line parallel to OY cuts (l) in not more than two points. The formula can easily be shown to be justified for more general contours. We first take the case when (σ) is bounded by curves (1) and (2) and by two straight segments parallel to OY (Fig. 62). We find on going through the same argument that

$$\sigma = -\left[\int_{(1)} y\,\mathrm{d}x + \int_{(2)} y\,\mathrm{d}x\right].$$

But x is constant over CD and BA and $\mathrm{d}x = 0$, so that $\int y\,\mathrm{d}x$ is zero over these sections. We can add these integrals with the minus sign to the right-hand side and thus obtain (13) for the case in questions. When (σ) has the more general type of contour (l) of Fig. 63, we proceed as follows. We draw straight lines parallel to OY so as to divide (σ) into a finite number of parts, to each of which (13) is applicable; addition of the expressions obtained gives us on the left-hand side the area σ of the total domain, and the integral over (l) on the right,

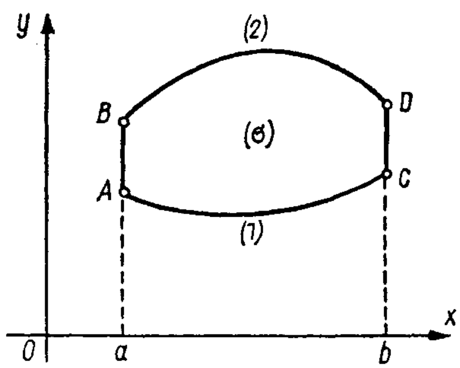

FIG. 62

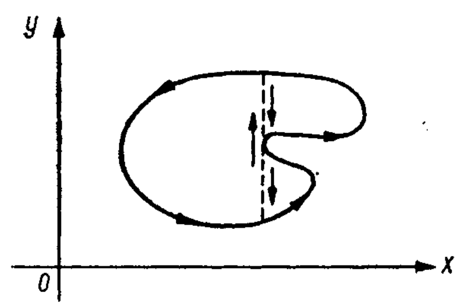

FIG. 63

the integrals over the additional contours being zero as above. Thus (13) is valid of the domain, and (14) and (15) may be justified for contours of general type in the same way.

In the case of the ellipse

$$x = a\cos t; \quad y = b\sin t \qquad (0 \leqslant t \leqslant 2\pi)$$

(15) gives:

$$\sigma = \frac{1}{2}\int_0^{2\pi} (a\cos t\cdot b\cos t + b\sin t\cdot a\sin t)\,\mathrm{d}t = \frac{1}{2}ab\int_0^{2\pi}\mathrm{d}t = \pi ab.$$

It is essential in the above formulae for an area for the circuit round (l) to be taken counter-clockwise, or more strictly, to be taken in the same direction as the $90°$ rotation needed to make OX coincide in direction with OY. If OY is directed downwards instead of upwards, the formulae remain in force provided a clockwise circuit is taken round (l). We shall always assume the above condition for the direction round a closed contour in a plane.

69. Green's formula. We now establish a fundamental formula relating an integral over a closed surface to a line integral over the contour of the surface. We start with the case when the surface is a plane domain, when we obtain what is usually known as *Green's formula.*

We use (7) of [56] to evaluate the double integral:

$$\iint\limits_{(\sigma)} \frac{\partial P(x,y)}{\partial y}\,d\sigma,$$

where $P(x, y)$ is a function of (x, y).

We integrate first with respect to y and assume that the contour (l) of (σ) is intercepted at only two points by lines parallel to OY (Fig. 61); this gives us:

$$\iint\limits_{(\sigma)} \frac{\partial P}{\partial y}\,d\sigma = \iint\limits_{(\sigma)} \frac{\partial P}{\partial y}\,dx\,dy = \int\limits_a^b dx \int\limits_{y_1}^{y_2} \frac{\partial P}{\partial y}\,dy =$$

$$= \int\limits_a^b \left[P(x, y_2) - P(x, y_1)\right] dx\,.$$

On the other hand, the integrals:

$$\int\limits_a^b P(x, y_1)\,dx, \quad \int\limits_b^a P(x, y_2)\,dx$$

are the same as the line integrals:

$$\int P(x, y)\,dx,$$

taken respectively over parts (1) and (2) of contour (l) from the point $x = a$ to $x = b$.

We get on changing the direction of integration of the second integral:

$$\int\limits_a^b P(x, y_2)\,dx = -\int\limits_b^a P(x, y_2)dx = -\int\limits_{(2)\,b}^{a} P(x, y)\,dx,$$

whence

$$\iint_{(\sigma)} \frac{\partial P}{\partial y}\, d\sigma = -\int_{(2)b}^{a} P(x, y)\, dx - \int_{(1)a}^{b} P(x, y)\, dx,$$

or

$$\iint_{(\sigma)} \frac{\partial P}{\partial y}\, d\sigma = -\int_{(l)} P\, dx. \tag{16}$$

where the circuit round (l) must be taken counter-clockwise (Fig. 61)

We find in the same way the integral:

$$\iint_{(\sigma)} \frac{\partial Q(x, y)}{\partial x}\, d\sigma,$$

where Q is another function of (x, y). Assuming for simplicity that (l) is intersected in only two points by lines parallel to OX, we have:

$$\iint_{(\sigma)} \frac{\partial Q}{\partial x}\, d\sigma = \iint_{(\sigma)} \frac{\partial Q}{\partial x}\, dx\, dy = \int_{\alpha}^{\beta} dx \int_{x_1}^{x_2} \frac{\partial Q}{\partial x}\, dx =$$

$$= \int_{\alpha}^{\beta} [Q(x_2, y) - Q(x_1, y)]\, dy,$$

and this can be similarly reduced to a line integral over a closed contour:

$$\iint_{(\sigma)} \frac{\partial Q}{\partial x}\, d\sigma = \int_{(l)} Q\, dy. \tag{17}$$

Subtraction of (16) from (17) gives us Green's formula:

$$\iint_{(\sigma)} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) d\sigma = \int_{(l)} P\, dx + Q\, dy. \tag{18}$$

We deduced (16) on the assumption that lines parallel to OY cut (l) in not more than two points. The same arguments as in the previous article can be used to show that the formula is justified for any type of contour. Similar remarks apply as regards (17) and (18).

The discussion is applicable to the case when (σ) is bounded by several curves (Fig. 64). The integration on the right-hand side of (18) is now over all the curves that bound the domain, in a counter-clockwise direction for the outer contour (with the axes directed as shown) but clockwise for the inner contours, i.e. so that the domain (σ) lies on the left over each contour.

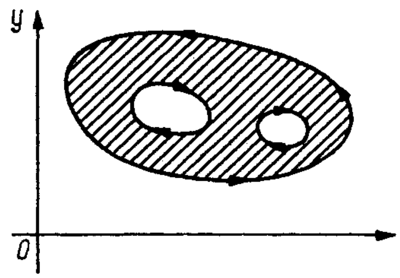

Fig. 64

We note that Green's formula (18) can be written in another way. Let t be the tangent to l, taken in the same direction, and let n be the normal to l, taken outwards from σ. The direction of t is found by turning n counter-clockwise through a right angle, so that we have for the angles formed by t and n with the axes: $(t, X) = \pi + (n, Y)$ and $(t, Y) = (n, X)$. If ds is an elementary arc of the curve, $dx = ds \cdot \cos(t, X)$ and $dy = ds \cdot \cos(t, Y)$, so that $dx = -ds \cdot \cos(n, Y)$ and $dy = ds \cdot \cos(n, X)$. If we substitute these in (18) and replace P by $-Q$) and Q by P, we get:

$$\iint_{(\sigma)} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) d\sigma = \int_{(l)} [P \cos(n, X) + Q \cos(x, Y)]\, ds.$$

Green's formula in this form amounts to Ostrogradskii's formula for a plane.

70. Stokes' formula. We now turn to the case of any non-closed surface (S) with contour l (Fig. 65). We retain the notation of [62], and assume that lines parallel to the z axis cut S in a single point. The projection of l on the XY plane gives the contour (λ) of the domain (σ_{xy}). We take the counter-clockwise circuit round (λ) as positive, and similarly for (l). The direction n of the normal to S is taken so that it forms an acute angle with axis OZ and $\cos(n, Z) > 0$. With this, the lower sign must be taken in expressions (24) of [62], which give:

$$p \cos(n, Z) = -\cos(n, X); \quad q \cos(n, Z) = -\cos(n, X), \tag{19}$$

whilst (26) of [62] can be written as:

$$d\sigma_{xy} = \cos(n, Z)\, dS. \tag{20}$$

Let $P(x, y, z)$ be a given function in the neighbourhood of surface (S) which is continuous and has continuous first order derivatives. We consider the integral:

$$\int_{(l)} P\,(x, y, z)\,\mathrm{d}x\,.$$

Points of (l) lie on (S) and therefore satisfy its equation: $z = f(x, y)$, and we can replace the z under the integral sign by $f(x, y)$. The integrand, $P[x, y, f(x, y)]$, now contains only x and y. The coordinates (x, y) of a variable point of (λ) are the same as for the corresponding point of (l), and integration round (l) can be replaced by integration round (λ):

$$\int_{(l)} P\,(x, y, z)\,\mathrm{d}x =$$

$$= \int_{(\lambda)} P\left[x, y, f\,(x, y)\right]\mathrm{d}x\,.$$

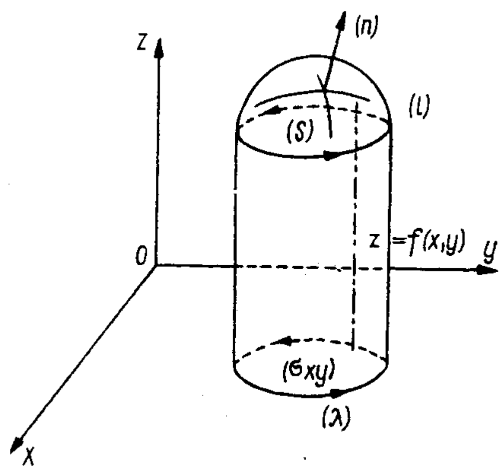

FIG. 65

We apply Green's formula (18) to the integral on the right, with $P = P[x, y, f(x, y)]$, $Q = 0$, and (l) becoming (λ) in this case. We first find $\partial P/\partial y$ by differentiating P both directly with respect to y and via the third argument z, which we replaced by $f(x, y)$:

$$\frac{\partial P}{\partial y} = \frac{\partial P\,(x, y, z)}{\partial y} + \frac{\partial P\,(x, y, z)}{\partial z} \cdot \frac{\partial f\,(x, y)}{\partial y}\,,$$

where $f(x, y)$ is to be understood for z in the expression for P. We now have from (18):

$$\int_{(l)} P\,(x, y, z)\,\mathrm{d}x = \int_{(\lambda)} P\left[x, y, f\,(x, y)\right]\mathrm{d}x =$$

$$= -\iint_{(\sigma_{xy})} \left[\frac{\partial P(x, y, z)}{\partial y} + \frac{\partial P(x, y, z)}{\partial z} \cdot \frac{\partial f(x, y)}{\partial y}\right]\mathrm{d}\sigma_{xy}.$$

On expressing $\mathrm{d}\sigma_{xy}$ in terms of the element $\mathrm{d}S$ of S in accordance with (20), we transform the double integral to an integral over (S) [63]:

$$\int_{(l)} P(x, y, z)\,\mathrm{d}x =$$
$$= -\iint_{(S)} \Big[\frac{\partial P(x, y, z)}{\partial y} + \frac{\partial P(x, y, z)}{\partial z} \cdot \frac{\partial f(x, y)}{\partial y}\Big] \cos(n, Z)\,\mathrm{d}S,$$

and obtain finally, on using the second of expressions (19):

$$\int_{(l)} P\mathrm{d}x = \iint_{(S)} \Big[\frac{\partial P}{\partial z}\cos(n, Y) - \frac{\partial P}{\partial y}\cos(n, Z)\Big]\,\mathrm{d}S. \tag{21}$$

If we are given two other functions, $Q(x, y, z)$ and $R(x, y, z)$, in the neighbourhood of (S), we can change the variables x, y, z cyclically to get two analogous expressions:

$$\int_{(l)} Q\,\mathrm{d}y = \iint_{(S)} \Big[\frac{\partial Q}{\partial x}\cos(n, Z) - \frac{\partial Q}{\partial z}\cos(n, X)\Big]\,\mathrm{d}S$$
$$\int_{(l)} R\,\mathrm{d}z = \iint_{(S)} \Big[\frac{\partial R}{\partial y}\cos(n, X) - \frac{\partial R}{\partial x}\cos(n, Y)\Big]\,\mathrm{d}S.$$

We arrive at Stokes' formula by adding these three expressions:

$$\int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z = \iint_{(S)} \Big[\Big(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\Big)\cos(n, X) +$$
$$+ \Big(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\Big)\cos(n, Y) + \Big(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\Big)\cos(n, Z)\Big]\,\mathrm{d}S. \tag{22}$$

The formula relates the line integral round the contour of a surface to the integral over the surface itself, and bears some similarity to the Ostrogradskii formula of [63], which relates the integral over the surface of a three-dimensional domain to the integral over the domain itself. Green's formula is the special case of Stokes' formula when (S) is a plane domain in the XY plane; (l) now becomes a closed curve in the XY plane and $\mathrm{d}z = 0$; (n) is along the z axis, so that $\cos(n, X) = \cos(n, Y) = 0$ and $\cos(n, Z) = 1$. We obtain (18), on making these substitutions in (22).

The same assumptions are made regarding the cosines in (22) as when deducing Ostrogradskii's formula [63].

We arrived at (21) on the assumption that a straight line parallel to the z axis cuts (S) in a single point. If this is not the case, we divide

(S) with the aid of supplementary curves so that our condition is satisfied over each portion of (S), and (21) is applied to each portion in turn; addition of the results gives us the integral over (l) on the left, since the integrals are taken twice in opposite directions over the supplementary contours and cancel out; the double integral over the total surface (S) is obtained on the right. Hence, (21) is justified in the general case. The same arguments apply for the general formula (22). It is only necessary to observe the following condition here for the circuit round (l) and the direction (n) of the normal: *an observer travelling round* (l), *with his height along the direction of the normal* (n), *must have the surface* (S) *on his left*. This rule is bound up with the choice of the system of coordinates shown in Fig. 65. In this system, the observer with his height along OZ, sees OX become OY by a counter-clockwise rotation through an angle of $\pi/2$. If the rotation were clockwise, "on his left" would have to be replaced by "on his right" in the above rule.

If the notation of [64] is used for the surface integral, (22) may be written in the form

$$\int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z =$$
$$= \iint_{(S)} \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathrm{d}y\,\mathrm{d}z + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathrm{d}z\,\mathrm{d}x +$$
$$+ \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x\,\mathrm{d}y. \tag{23}$$

The side of the surface (S) and the direction (n) are defined in accordance with the rule above.

71. Independence of a line integral on the path in a plane. The examples in [67] showed that the value of a line integral is sometimes independent of the path of integration, though this is not always the case. We now use Green's and Stokes' formulae to show the conditions for independence of the integral on the path, and start with the case of a plane curve and the condition for independence of

$$\int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y$$

on the path. We join the points (A) and (B) with curves (1) and (2) (Fig. 66), when we must have

$$(1) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y = (2) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y, \tag{24}$$

or, on using property II of [66]:

$$(1) \quad \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y - (2) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y = 0,$$

$$(1) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y + (2) \int_{(B)}^{(A)} P\,\mathrm{d}x + Q\,\mathrm{d}y = \int_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y = 0, \tag{25}$$

where (l) is a closed contour, formed by curve (1) taken from (A) to (B) and curve (2) taken from (B) to (A). We see from this that the integral over any closed contour (l) must be zero, in view of the arbitrariness of points A and B. Conversely, if the integral over a closed contour (l) is zero, the integral over (1) is equal to the integral over (2), since (24) follows conversely from (25). If curves (1) and (2), that join A and B, intersect, we join A and B with another curve (3) that cuts neither (1) nor (2), and use the equations

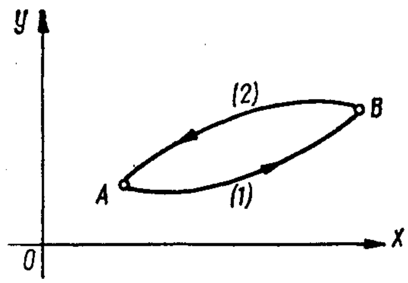

FIG. 66

$$(1) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y = (3) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y$$

$$(2) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y = (3) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y$$

to obtain:

$$(1) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y = (2) \int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y.$$

It follows that *the condition for independence of the integral on the path is the same as the condition that the integral over any closed contour (l) vanishes.*

If the latter condition is satisfied, we obtain from (18):

$$\iint_{(\sigma)} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \mathrm{d}\sigma = 0\,, \tag{26}$$

where the domain of integration (σ) is *entirely arbitrary.*

We now prove the consequence, that

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0 \tag{27}$$

is an *identity*, i.e. true for all x and y.

Suppose that at a point $C(a, b)$,

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = f(x, y)$$

differs from zero and is, say, positive. By the assumed continuity of $\partial P/\partial y$ and $\partial Q/\partial x$, $f(x, y)$ must be positive over some small circle (σ_0) with centre C. We form the integral

$$\iint\limits_{(\sigma_0)} f(x, y)\,\mathrm{d}\sigma = \iint\limits_{(\sigma_0)} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}\sigma,$$

and apply the mean value theorem [61] to it:

$$\iint\limits_{(\sigma_0)} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}\sigma = f(\xi, \eta)\,\sigma,$$

where (ξ, η) is a point of (σ_0), so that $f(\xi, \eta) > 0$; hence it follows that

$$\iint\limits_{(\sigma_0)} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}\sigma > 0,$$

which contradicts the fact that integral (26) vanishes for any choice of (σ). Condition (27) is thus necessary for the independence of the integral on the path. It is easily shown to be sufficient, since it follows from this, by (18), that $\int\limits_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y$ vanishes for any closed contour, which is the same as the independence of the integral on the path.

Condition (27) *is thus necessary and sufficient for the integral*

$$\int\limits_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y \tag{28}$$

to be independent of the path of integration, and to be a function of the coordinates of points A and B only.

If this condition is satisfied and we fix the point $A(x_0, y_0)$ whilst only the point $B(x, y)$ varies, integral (28) is a function of (x, y), i.e. a function of the point B:

$$\int\limits_{(x_0, y_0)}^{(x, y)} P\,\mathrm{d}x + Q\,\mathrm{d}y = U(x, y). \tag{29}$$

We consider the properties of this function. On keeping y constant and giving x an increment Δx, we get:

$$U(x+\Delta x, y) - U(x, y) = \int_{(x_0, y_0)}^{(x+\Delta x, y)} P\,dx + Q\,dy - \int_{(x_0, y_0)}^{(x, y)} P\,dx + Q\,dy.$$

Since the integral is independent of the path of integration, we can take the path for the first integral as made up of the curve AB (Fig. 67) as in the second integral, together with the straight line BB'. The integrals over AB cancel out, and we are left with:

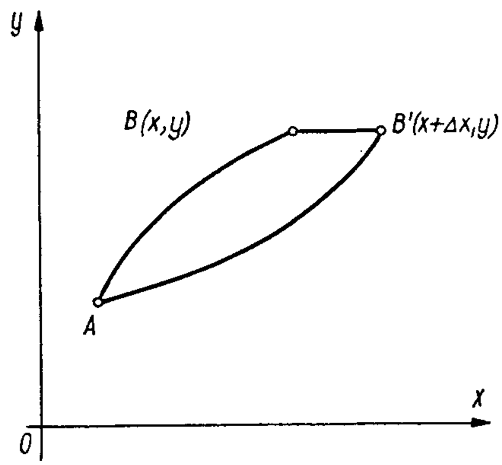

FIG. 67

$$U(x+\Delta x, y) - U(x, y) =$$

$$= \int_{(x, y)}^{(x+\Delta x, y)} P\,dx + Q\,dy =$$

$$= \int_{x}^{x+\Delta x} P(x, y)\,dx,$$

since y is constant on BB', and $dy = 0$. We apply the mean value theorem [I, 95]:

$$U(x+\Delta x, y) - U(x, y) = \Delta x P(x + \theta\Delta x, y) \qquad (0 < \theta < 1).$$

We divide by Δx and let $\Delta x \to 0$:

$$\frac{\partial U}{\partial x} = \lim_{\Delta x \to 0} P(x + \theta\Delta x, y) = P(x, y). \tag{30}$$

It can be shown in a similar manner that:

$$\frac{\partial U}{\partial y} = Q(x, y). \tag{31}$$

We find from (30) and (31) that [I, 68]:

$$dU = \frac{\partial U}{\partial x}\,dx + \frac{\partial U}{\partial y}\,dy = P\,dx + Q\,dy.$$

Hence, if condition (27) is satisfied, the integrand

$$P\,dx + Q\,dy \tag{32}$$

is the total differential of the function $U(x, y)$ defined by (29). It is easily seen that the most general expression for the function $U_1(x, y)$, whose total differential is (32), is given by

$$U_1(x, y) = U(x, y) + C, \tag{33}$$

where C is an arbitrary constant. We have, in fact:

$$\mathrm{d}U = P\,\mathrm{d}x + Q\,\mathrm{d}y,$$

$$\mathrm{d}U_1 = P\,\mathrm{d}x + Q\,\mathrm{d}y,$$

whence

$$\mathrm{d}(U_1 - U) = 0.$$

If the differential of a function is identically zero, its partial derivatives with respect to the independent variables must be zero, and the function itself is therefore a constant, i.e.

$$U_1 - U = C,$$

which is what we wished to prove.

We notice the obvious equality, valid when (27) is satisfied:

$$\int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y = \int_{(A)}^{(B)} \mathrm{d}U_1 = U_1(B) - U_1(A). \tag{34}$$

Conversely, let U_1 exist such that:

$$\mathrm{d}U_1 = P\,\mathrm{d}x + Q\,\mathrm{d}y. \tag{35}$$

We show that now:

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0,$$

and that U_1 is given by

$$U_1(x, y) = \int_{(x_0, y_0)}^{(x, y)} [P\,\mathrm{d}x + Q\,\mathrm{d}y] + C.$$

Equation (35) can be written as

$$P\,\mathrm{d}x + Q\,\mathrm{d}y = \frac{\partial U_1}{\partial x}\,\mathrm{d}x + \frac{\partial U_1}{\partial y}\,\mathrm{d}y,$$

and since the differentials $\mathrm{d}x$ and $\mathrm{d}y$ of the independent variables are completely arbitrary [I, 68], it can only be satisfied when the coefficients of $\mathrm{d}x$ and $\mathrm{d}y$ are the same on both sides, i.e.

$$P = \frac{\partial U_1}{\partial x}; \quad Q = \frac{\partial U_1}{\partial y},$$

whence it is clear that:

$$\frac{\partial P}{\partial y} = \frac{\partial^2 U_1}{\partial x\, \partial y} = \frac{\partial^2 U_1}{\partial y\, \partial x} = \frac{\partial Q}{\partial x}\,.$$

Hence (27) is fulfilled, and the integral

$$U(x, y) = \int_{(x_0, y_0)}^{(x, y)} P\,\mathrm{d}x + Q\,\mathrm{d}y$$

is now, by the previous discussion, only dependent on (x, y) and has the property that:

$$\mathrm{d}U = P\,\mathrm{d}x + Q\,\mathrm{d}y = \mathrm{d}U_1,$$

whence it follows that:

$$U_1 = U + C,$$

which was required to be proved. We can thus say: *a necessary and sufficient condition for $P\,\mathrm{d}x + Q\,\mathrm{d}y$ to be the total differential of a function U_1 is given by the identity:*

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}\,,$$

U_1 being then given by:

$$U_1(x, y) = \int_{(x_0, y_0)}^{(x, y)} P\,\mathrm{d}x + Q\,\mathrm{d}y + C. \tag{36}$$

72. Multiply connected domains. The proof of the necessity and sufficiency of (27) for the line integral

$$\int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y$$

to be independent of the path is based on the following two essential assumptions:

(1) Functions P and Q and their first order partial derivatives must be continuous in the domain of variation of (x, y).

(2) Whatever closed contour (l) is drawn in the domain, the part of the plane inside (l) belongs to the domain where (27) and the continuity condition are satisfied.

The first condition is important because the functions mentioned come under the integral sign in the proof. The second is needed for the application of Green's formula, i.e. for the transformation from a line to a double integral. It is equivalent to the fact that any

closed contour traced in the domain can be contracted continuously to a point without moving out of the domain, or to put the matter more simply, the condition amounts to the domain having no gaps.

We now suppose that P, Q and their derivatives are continuous and that (27) is satisfied in a domain (σ) with two gaps or holes (Fig. 68). If a closed contour (l_0) with no holes in its interior is taken in the domain, Green's formula (18) is applicable to the contour and its interior, and by (27), the integral round (l_0) must be zero. We now take a closed contour (l_1) containing hole (I). We cannot apply (18) here, and in general, integral (28) round (l_1) does not vanish. We show that the value of this integral is independent of the form of (l_1), the only important factor being that the contour encircles the single hole (I).

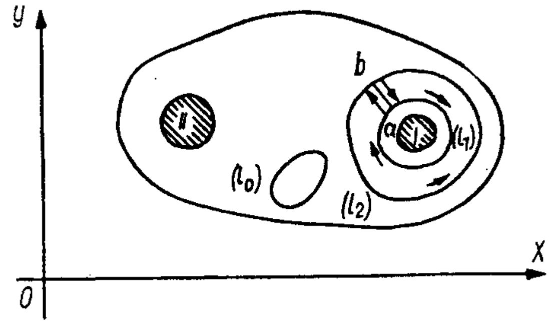

FIG. 68

Let (l_1) and (l_2) be two contours encircling (I). We have to show that integral (28) round (l_1) is the same as round (l_2). We join (l_1) and (l_2) by means of the extra contour (ab), so that the three together form a closed contour without a hole in the domain; the circuit round the total contour should be as shown by the arrows. We can apply (18) to the total contour and the integral round it must vanish, by (27):

$$\int_{\circlearrowright(l_1)} + \int_{(ba)} + \int_{\circlearrowleft(l_2)} + \int_{(ab)} = 0.$$

The integrals over (ba) and (ab) are taken in opposite directions and cancel out, whilst the integration is clockwise round (l_1) and counter-clockwise round (l_2). We can change the direction of integration round (l_1) providing we change the sign in front of the integral, so that we get:

$$\int_{\circlearrowleft(l_2)} - \int_{\circlearrowleft(l_1)} = 0$$

or finally:

$$\int_{\circlearrowleft(l_1)} P\,\mathrm{d}x + Q\,\mathrm{d}y = \int_{\circlearrowleft(l_2)} P\,\mathrm{d}x + Q\,\mathrm{d}y,$$

i.e. the integrals round (l_1) and (l_2) are equal, provided they are both taken counter-clockwise in the usual way. It follows that *hole* (I)

corresponds to a definite constant ω_1, *equal to the value of integral* (28) *taken over any closed contour encircling* (I). *Similarly, hole* (II) *corresponds to another constant* ω_2.

If we make two slits (ab) and (cd) in a domain (D) from the holes to the outer edge (Fig. 69), a new domain is obtained without holes, and by (27), we can form the single-valued function in the domain:

$$U(x, y) = \int_{(x_0, y_0)}^{(x, y)} P\,\mathrm{d}x + Q\,\mathrm{d}y;$$

but, by the above, the value of the function changes by a constant ω_1 on crossing (ab), and by ω_2 on crossing (cd). If we remove the slits and return to the original domain (D), *the function* $U(x, y)$ *is many-valued*. Circuits round the holes imply adding ω_1 and ω_2 to the function, i.e. $U(x, y)$ *includes an arbitrary term of the form* $m_1\omega_1 + m_2\omega_2$, where m_1 and m_2 *are arbitrary integers*. Our arguments are obviously all applicable to any number of holes in a domain, including the case when a hole consists of a single point. *The degree of connectivity of a domain* (D) is defined as the number of holes plus one; the domain with holes is said to be *multiply connected*. Constants ω_1 and ω_2 are called the *circulations* of $P\,\mathrm{d}x + Q\,\mathrm{d}y$ or the *cyclical constants* of $U(x, y)$.

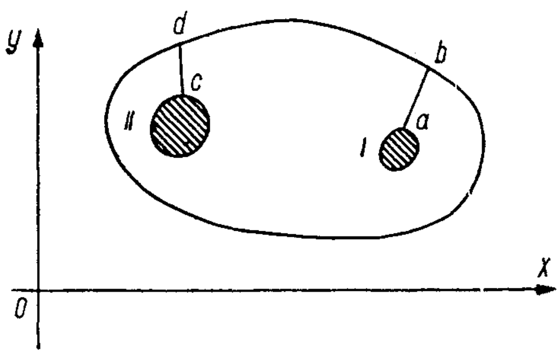

FIG. 69

Example. We take the function

$$\varphi = \operatorname{arc\,tan} \frac{y}{x},$$

given in a domain (D) bounded by two concentric circles with centre at the origin.

We define P and Q by:

$$\left.\begin{aligned} P &= \frac{\partial\varphi}{\partial x} = -\frac{y}{x^2 + y^2} \\ Q &= \frac{\partial\varphi}{\partial y} = \quad \frac{x}{x^2 + y^2} \end{aligned}\right\} \tag{37}$$

These functions are continuous with their derivatives in (D) and are easily shown to satisfy (27). We take the line integral

$$\int\limits_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y = \int\limits_{(l)} \frac{-y\,\mathrm{d}x + x\,\mathrm{d}y}{x^2 + y^2}$$

and integrate round circle (l_1) with centre at the origin and radius a. We get by substituting $x = a\cos\varphi$; $y = a\sin\varphi$:

$$\int\limits_{(l_1)} \frac{-y\,\mathrm{d}x + x\,\mathrm{d}y}{x^2 + y^2} = \int\limits_0^{2\pi} \mathrm{d}\varphi = 2\pi.$$

The domain (D) has one hole in this case, and the cyclical constant $\omega_1 = 2\pi$. The function $U_1(x, y)$ is the polar angle:

$$U_1(x, y) = \int P\,\mathrm{d}x + Q\,\mathrm{d}y = \int \frac{\partial\varphi}{\partial x}\,\mathrm{d}x + \frac{\partial\varphi}{\partial y}\,\mathrm{d}y = \varphi$$

and it increases by 2π on a circuit round the hole. We note that the radius of the inner circle can be taken as zero, when we have a point-hole, i.e. we exclude the origin $(0, 0)$. By (37), P and Q take the indeterminate form $0/0$ at the origin.

73. Independence of a line integral on the path in space. Just as in the case of a plane, the condition for a line integral to be independent of the path in space amounts to the vanishing of the integral round any closed contour. We take the integral

$$\int\limits_{(l)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z. \tag{38}$$

We can show as above, on using Stokes' formula (22), that *the necessary and sufficient conditions for the independence of integral* (38) *on the path consist of the three identities*:

$$\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} = 0; \quad \frac{\partial P}{\partial z} - \frac{\partial R}{\partial z} = 0; \quad \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0. \tag{39}$$

If these conditions are satisfied, we can form a function U of points (x, y, z):

$$U(x, y. z) = \int\limits_{(x_0, y_0, z_0)}^{(x, y, z)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z, \tag{40}$$

where it can be shown by a method similar to the previous one that

$$\frac{\partial U}{\partial x} = P; \quad \frac{\partial U}{\partial y} = Q; \quad \frac{\partial U}{\partial z} = R \tag{41}$$

$$P\,\mathrm{d}z + Q\,\mathrm{d}y + R\,\mathrm{d}z = \mathrm{d}U \tag{42}$$

$$\int_{(A)}^{(B)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z = U(B) - U(A). \tag{43}$$

Furthermore, conditions (39) are necessary and sufficient for the expression $P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z$ to be the total differential of a function U_1, where U_1 is given by:

$$U_1 = \int_{(x_0,\, y_0,\, z_0)}^{(x,\, y,\, z)} P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z + C,$$

where C is an arbitrary constant.

Multiply connected domains in space have certain special features. We take the example of the domain (D) formed by the interior of a sphere, from which two tubular portions (I) and (II) are cut out, with their ends on the surface of the sphere, as shown in Fig. 70. If we take a closed contour (l_1) round tube (I), a surface contained in (D) cannot be shrunk on to it, so that even if conditions (39) are satisfied in (D), Stokes' formula cannot be applied to (l_1), and integral (38) round (l_1) will in general differ from zero. The non-zero value of the integral will be independent of the shape of (l_1), however; it gives the cyclical constant ω_1 for tube (I) and is the same for any closed contour in (D) that encircles (I). Similarly, we get a second cyclical constant ω_2 for tube (II). The function $U(x, y, z)$ given by (40) is many-valued in this case and includes the undefined term $m_1\,\omega_1 + m_2\,\omega_2$, where m_1 and m_2 are arbitrary integers.

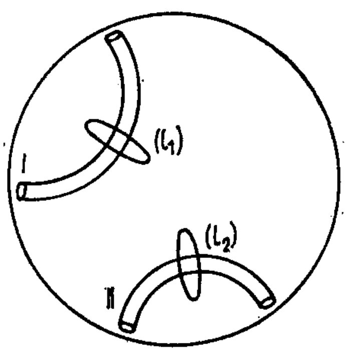

FIG. 70

We notice that if (D) consists of the space between two concentric spheres, and conditions (39) are satisfied in (D), there are no cyclical constants and (40) is a single-valued function. This is obvious geometrically from the fact that a surface lying in (D) can be shrunk to any closed contour in (D), so that Stokes' formula (22) can be applied to the contour; the vanishing of the integral round the contour then follows from (39).

Example. We consider the angle φ appearing in systems of cylindrical and spherical coordinates:

$$\varphi = \operatorname{arc} \tan \frac{y}{x},$$

and we use (37) to define P and Q. These latter take the indeterminate form 0/0 all along the z axis. When we consider the line integral in space

$$\int_{(l)} P \, dx + Q \, dy = \int_{(l)} \frac{-y \, dx + x \, dy}{x^2 + y^2}$$

we have to exclude the tube running along the z axis, the cyclical constant 2π being obtained as the value of the integral written round any closed contour encircling this tube.

74. Steady-state flow of fluids. Let $\mathbf{v}(x, y)$ be the velocity vector of the steady flow of an incompressible fluid in a plane, and let $u(x, y)$, $v(x, y)$ be the projections of the vector on the axes. We saw in Example 3 [67] that the absence of sources implies the vanishing of the integral

$$\int_{(l)} -v \, dx + u \, dy \tag{44}$$

over any closed contour, which is equivalent to the integral being independent of the path; the necessary and sufficient condition for this is, by (27):

$$\frac{\partial(-v)}{\partial y} = \frac{\partial u}{\partial x}, \quad \text{or} \quad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \tag{45}$$

which characterizes an incompressible fluid. When (45) is satisfied, the expression

$$-v \, dx + u \, dy$$

is the total differential of some function $\psi(M)$, given by the relationship

$$\psi(B) - \psi(A) = \int_{(A)}^{(B)} -v \, dx + u \, dy. \tag{46}$$

Function $\psi(M)$ is called the *stream function* and has a simple physical significance: *the quantity of fluid flowing per unit time past an arbitrary contour starting at A and finishing at B is given by* $\psi(B) - \psi(A)$. This follows immediately from formula (10) for the quantity of fluid [67].

If certain points of the domain are sources, we can exclude these points and get a domain with holes in which (45) is valid. The cyclical constant for a hole is equal to integral (44) over a contour encircling the hole, and evidently gives the quantity of fluid q produced by the corresponding source in unit time. The function $\psi(M)$ is now many-valued. If $q < 0$, the source is negative (a sink).

We consider, in addition to (44), the integral

$$\int_{(l)} u \, dx + v \, dy, \tag{47}$$

the value of which is usually called the speed of circulation along contour (l). We take the speed of circulation over a closed contour as zero, i.e. (47) is independent of the path. This is the same as saying that the flow has no vortices. With this assumption, a function $\varphi(M)$ exists:

$$\varphi = \int_{(x_0, y_0)}^{(x, y)} u\,dx + v\,dy, \tag{48}$$

such that the projections u and v of vector $\mathbf{v}$ are its partial derivatives:

$$u = \frac{\partial \varphi}{\partial x}; \quad v = \frac{\partial \varphi}{\partial y}. \tag{49}$$

The function φ is called the velocity potential. If we have independence of integral (48) on the path in a multiply connected domain (a domain with holes), the velocity potential φ will in general be a many-valued function and the cyclical constants of (48) with respect to each hole will give the strength of the corresponding vortex.

Equation (46) gives us [71]:

$$-v = \frac{\partial \psi}{\partial x}; \quad u = \frac{\partial \psi}{\partial y}.$$

On comparing these equations with (49), we get two equations relating the velocity potential φ and the stream function ψ:

$$\frac{\partial \varphi}{\partial x} = \frac{\partial \psi}{\partial y}; \quad \frac{\partial \varphi}{\partial y} = -\frac{\partial \psi}{\partial x}. \tag{50}$$

These are the Cauchy-Riemann equations; they are of fundamental importance in the theory of functions of a complex variable and their hydrodynamical significance, established above, serves as a basis for the wide applications of the theory of functions of complex variables to plane problems of hydrodynamics.

In the case of steady-state motion in space, the velocity vector $\mathbf{v}(x, y, z)$ has three components: $u(x, y, z)$, $v(x, y, z)$, $w(x, y, z)$, whilst we have to consider, instead of (48),

$$\int_{(l)} u\,dx + v\,dy + w\,dz,$$

the conditions for this to be independent of the path being:

$$\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} = 0; \quad \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} = 0; \quad \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0,$$

if these conditions are fulfilled, there exists the velocity potential

$$\varphi = \int_{(x_0, y_0, z_0)}^{(x, y, z)} u\,dx + v\,dy + w\,dz,$$

where

$$u = \frac{\partial \varphi}{\partial x}; \quad v = \frac{\partial \varphi}{\partial y}; \quad w = \frac{\partial \varphi}{\partial z}.$$

The generalization of condition (45) for incompressibility to the case of space will be found in the next chapter.

75. Integrating factors. If

$$P\,\mathrm{d}x + Q\,\mathrm{d}y \tag{51}$$

is not a total differential, i.e.

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} \neq 0,$$

it can be shown that a function μ can always be found which yields a total differential when multiplied by (51), i.e.

$$\mu\,(P\,\mathrm{d}x + Q\,\mathrm{d}y) = \mathrm{d}U. \tag{52}$$

The function μ is called an *integrating factor* of (51).

By [71], a necessary and sufficient condition for μ to be an integrating factor of (51) is for the equation

$$\frac{\partial\,(\mu P)}{\partial y} - \frac{\partial\,(\mu Q)}{\partial x} = 0, \tag{53}$$

to be satisfied; this can be rewritten as

$$P\,\frac{\partial \mu}{\partial y} - Q\,\frac{\partial \mu}{\partial x} + \mu\left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x}\right) = 0. \tag{54}$$

which can be looked on as an equation for μ. This is in general difficult to use, however, since, being a partial differential equation, its integration is likely to be more awkward than would be the case with an ordinary equation.

If (51) is a total differential, the equation

$$P\,\mathrm{d}x + Q\,\mathrm{d}y = 0 \tag{55}$$

is called an *exact differential equation.*

It may be integrated directly. Let U be the function for which

$$\mathrm{d}U = P\,\mathrm{d}x + Q\,\mathrm{d}y.$$

Given our assumptions, which are equivalent to condition (27), this function can always be found from (29). Equation (55) is the same as $\mathrm{d}U = 0$, i.e.

$$U = C, \tag{56}$$

this being the general solution of the given equation (55).

Now suppose that (51) is not a total differential. By the existence theorem [51], (55) will always have a general solution, which may be written in the form:

$$F(x, y) = C.$$

The function $F(x, y)$ must satisfy the relationship

$$\frac{\partial F(x, y)}{\partial x} + \frac{\partial F(x, y)}{\partial y} \cdot \frac{dy}{dx} = 0,$$

where dy/dx must be replaced, using (55), by $(-P/Q)$, i.e. we must have the identity [7]:

$$\frac{\frac{\partial F}{\partial x}}{P} = \frac{\frac{\partial F}{\partial y}}{Q}.$$

If μ is the common value of these ratios, we have

$$\frac{\partial F}{\partial x} = \mu P; \qquad \frac{\partial F}{\partial y} = \mu Q,$$

i.e. μ is an integrating factor of (51).

The above shows that *every expression* $P\,dx + Q\,dy$ *has an integrating factor.*

If we find an integrating factor of (51), and from this the function F, the general solution of (55) can be written at once as

$$F = C.$$

Examples. **1.** The differential equation of a stream line in steady state plane fluid flow is [52]

$$\frac{dx}{u} = \frac{dy}{v} \quad \text{or} \quad -v\,dx + u\,dy = 0, \tag{57}$$

where u and v are the projections of the velocity $\mathbf{v}$ on the axes. If the fluid is incompressible, we have:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0,$$

which shows that

$$-v\,dx + u\,dy \tag{58}$$

is a total differential; we saw in [74], in fact, that

$$-v\,dx + u\,dy = d\psi.$$

where ψ is the stream function; the equation of a stream line is

$$\psi = C,$$

which is in fact the general solution of the equation:

$$-v\,dx + u\,dy = 0.$$

2. We discussed elementary heat processes in Example 4 of [67] and gave expressions for the infinitesimal quantity of heat obtained in the process as a function of the infinitesimal changes in pressure p, volume v, and temperature T.

We have three expressions for dQ, depending on which pair of p, v, T is taken for the independent variables:

$$dQ = \begin{cases} c_v\, dT + c_1\, dv & (T,\ v\text{ — independent variables}) \\ c_p\, dT + c_2\, dp & (T,\ p\text{ — } \quad ,, \qquad ,, \quad) \\ P\, dp + V\, dv & (p,\ v\text{ — } \quad ,, \qquad ,, \quad). \end{cases} \tag{59}$$

The fundamentally important quantities c_v and c_p are called the specific heats of the body at constant volume and constant pressure respectively.

On expressing one of the independent variables of (59) in terms of the other two, we get a series of relationships between the coefficients.

For instance, let T and v be the independent variables in the equation

$$c_p\, dT + c_2\, dp = c_v\, dT + c_1\, dv \tag{60}$$

We write:

$$dp = \frac{\partial p}{\partial T}\, dT + \frac{\partial p}{\partial v}\, dv.$$

and by substituting this for dp in (60) and equating coefficients of dT and dv, we get:

$$c_v = c_p + c_2 \frac{\partial p}{\partial T} \tag{61}$$

$$c_v = c_2 \frac{\partial p}{\partial v}\ . \tag{62}$$

In the same way, we use the equation·

$$c_v\, dT + c_1\, dv = P\, dp + V\, dv$$

to obtain:

$$c_v = P \frac{\partial p}{\partial T}\ , \tag{63}$$

$$c_1 = V + P \frac{\partial p}{\partial v}\ . \tag{64}$$

We have the equation of state, in the case of an ideal gas:

$$pv = RT,$$

whence the relationships follow:

$$\frac{\partial p}{\partial T} = \frac{R}{v},\ \frac{\partial p}{\partial v} = -\frac{p}{v};\quad \frac{\partial v}{\partial T} = \frac{R}{v};\quad \frac{\partial v}{\partial p} = -\frac{v}{p};\quad \frac{\partial T}{\partial p} = \frac{v}{R};\quad \frac{\partial T}{\partial v} = \frac{p}{R}\,,$$

whilst equations (61) to (64) give:

$$c_v = c_p + c_2 \frac{R}{v}\ ;\quad c_1 = -c_2 \frac{p}{v}\ ;\quad c_v = P\frac{R}{v}\ ;\quad c_1 = -P\frac{p}{v} + V\,. \tag{65}$$

These equations enable c_1, c_2, P, and V to be expressed in terms of the basic magnitudes c_v and c_p:

$$c_1 = (c_p - c_v)\frac{p}{R}; \quad c_2 = -(c_p - c_v)\frac{v}{R}; \quad P = c_v\frac{v}{R}; \quad V = c_p\frac{p}{R}. \tag{66}$$

The expression for dQ is not in general a total differential. But by the two basic principles of thermodynamics, it can be stated that:

I. The difference between dQ and the elementary work of the pressure, pdv, is a total differential:

$$\mathrm{d}Q - p\,\mathrm{d}v = \mathrm{d}U,$$

where U is called the *internal energy*.

II. The result of dividing dQ by the absolute temperature T is a total differential; in other words, $1/T$ is an integrating factor of dQ:

$$\frac{\mathrm{d}Q}{T} = \mathrm{d}S,$$

where S is called the *entropy*.

Rule I gives us, by the first of expressions (59):

$$\mathrm{d}U = \mathrm{d}Q - p\,\mathrm{d}v = c_v\,\mathrm{d}T + (c_1 - p)\,\mathrm{d}v,$$

whence

$$\left.\frac{\partial c_v}{\partial v}\right|_T = \left.\frac{\partial(c_1 - p)}{\partial T}\right|_v. \tag{67}$$

Subscripts T and v refer to the variables that are reckoned constant during the differentiations.

Similarly, rule II gives:

$$\mathrm{d}S = \frac{\mathrm{d}Q}{T} = \frac{c_v}{T}\,\mathrm{d}T + \frac{c_1}{T}\,\mathrm{d}v,$$

whence

$$\frac{1}{T}\left.\frac{\partial c_v}{\partial v}\right|_T = \left.\frac{\partial}{\partial T}\left(\frac{c_1}{T}\right)\right|_v = \frac{1}{T}\left.\frac{\partial c_1}{\partial T}\right|_v - \frac{c_1}{T^2},$$

or

$$\left.\frac{\partial c_v}{\partial v}\right|_T = \left.\frac{\partial c_1}{\partial T}\right|_v - \frac{c_1}{T}. \tag{68}$$

We find by comparing (67) and (68) that

$$\frac{\partial p}{\partial T} = \frac{c_1}{T}. \tag{69}$$

We obtain from this, on returning to the case of an ideal gas:

$$\frac{\partial p}{\partial T} = \frac{R}{v} = \frac{c_1}{T}; \qquad c_1 = \frac{RT}{v} = p. \tag{70}$$

On the other hand, (66) gives:

$$c_1 = p = (c_p - c_v)\frac{p}{R}, \quad \text{or} \quad c_p - c_v = R. \tag{71}$$

It may be taken from experimental data that:

III. The specific heat c_p of an ideal gas at constant pressure is a constant, so that $c_v = c_p - R$ is also a constant.

It follows from (71) that $c_p > c_v$, and if we write

$$\frac{c_p}{c_v} = k,$$

where $k > 1$, we easily get, on using (66) and (71):

$$c_1 = p; \quad c_2 = -v; \quad P = \frac{v}{k-1}; \quad V = p\frac{k}{k-1},$$

after which, (59) gives the following expressions for dQ, dU, dS:

$$dQ = \begin{cases} c_v\, dT + p\, dv \\ c_p\, dT - v\, dp \\ \dfrac{v\, dp + kp\, dv}{k-1} \end{cases} \tag{72}$$

$$dU = c_v\, dT \tag{73}$$

$$dS = c_v\frac{dT}{T} + \frac{p}{T}dv = c_v\frac{dT}{T} + R\frac{dv}{v}. \tag{74}$$

The temperature remains constant in isothermal processes, i.e. $dT = 0$ and

$$dQ = p\, dv,$$

i.e. all the absorbed heat goes into work done by the pressure and the total change in the quantity of absorbed heat on passing from volume v_1 to v_2 is

$$\int_{(v_1)}^{(v_2)} p\, dv.$$

The graph of a constant temperature process is called an *isothermal*.

An *adiabatic process* is one in which there is neither influx nor loss of heat, and it is characterized by the condition:

$$dQ = 0 \quad \text{or} \quad dS = 0; \quad S = \text{const.},$$

or constant entropy. We can find the entropy from (74):

$$S = c_v \log T + R \log v + C,$$

so that an adiabatic process is characterized by

$$c_v \log T + R \log v = \text{const.},$$

or, in non-logarithmic form:

$$T^{c_v} v^R = T^{c_v} v^{c_p - c_v} = \text{const.},$$

or alternatively, on raising to the power $1/c_v$:

$$Tv^{k-1} = \text{const.},$$

whence finally, inasmuch as $T = pv/R$:

$$pv^k = \text{const.} \tag{75}$$

Finally, we have at constant volume, when $\mathrm{d}v = 0$:

$$\mathrm{d}Q = c_v\,\mathrm{d}T; \quad \mathrm{d}Q = c_v\,(T_2 - T_1), \tag{76}$$

if the gas passes from temperature T_1 to T_2.

76. Exact differential equations in the case of three variables. We obtain on generalizing (55) for three variables:

$$P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z = 0, \tag{77}$$

where P, Q and R are given functions of (x, y, z). If conditions (39) are satisfied, the left-hand side of (77) is the total differential of a function $U(x, y, z)$, and the general solution of (77) becomes:

$$U(x, y, z) = C, \tag{78}$$

where C is an arbitrary constant. Geometrically, (78) gives a family of surfaces in space. Assuming that the left-hand side of (77) is not a total differential, an integrating factor is sought, i.e. a function $\mu(x, y, z)$ such that the left-hand side of the equation

$$\mu(P\,\mathrm{d}x + Q\,\mathrm{d}y + R\,\mathrm{d}z) = 0 \tag{79}$$

is in fact a total differential. Conditions (39) now give:

$$\frac{\partial(\mu R)}{\partial y} - \frac{\partial(\mu Q)}{\partial z} = 0; \qquad \frac{\partial(\mu P)}{\partial z} - \frac{\partial(\mu R)}{\partial x} = 0;$$
$$\frac{\partial(\mu Q)}{\partial x} - \frac{\partial(\mu P)}{\partial y} = 0,$$

which can be written as:

$$\left.\begin{aligned}
\mu\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) &= Q\frac{\partial\mu}{\partial z} - R\frac{\partial\mu}{\partial y}\,;\\
\mu\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) &= P\frac{\partial\mu}{\partial y} - Q\frac{\partial\mu}{\partial x}\,;\\
\mu\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) &= R\frac{\partial\mu}{\partial x} - P\frac{\partial\mu}{\partial z}.
\end{aligned}\right\} \tag{80}$$

If we multiply these equations by P, Q, R respectively, add, and cancel μ, we get a relationship between P, Q, R:

$$P\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + Q\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + R\left(\frac{\partial Q}{\partial y} - \frac{\partial P}{\partial y}\right) = 0. \tag{81}$$

Having assumed the existence of the integrating factor μ, we have thus arrived at the necessary condition (81) which must be satisfied by P, Q, R. We shall not dwell on the proof that the condition is also sufficient, i.e. *equation* (77) *does not always possess an integrating factor, the necessary and sufficient condition for the existence of such a factor being given by* (81). If μ exists, the left-hand side of (79) is the total differential of a function U, and (78) gives the general solution of equations (79) and (77). On the other hand, if condition (81) is not satisfied, (77) does not possess a general solution of the form (78). We sometimes refer to (81) as the condition for exact integrability of equation (77).

The geometrical significance of (77) and its general solution (78), when the latter exists, is as follows. The functions

$$P(x, y, z), \quad Q(x, y, z), \quad R(x, y, z)$$

may be looked on as defining the projections on the axes of a vector $\mathbf{v}(x, y, z)$ at every point. The system of differential equations

$$\frac{\mathrm{d}x}{P} = \frac{\mathrm{d}y}{Q} = \frac{\mathrm{d}z}{R}$$

defines a family of curves (L) in space, such that at every point of these the corresponding vector $\mathbf{v}$ is directed along the tangent. The stream lines of established flow in [52] played an exactly analogous role. Equation (77) is equivalent to the condition for an infinitesimal displacement with components $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$ to be perpendicular to vector $\mathbf{v}$, that is, (77) defines a plane element at every point perpendicular to $\mathbf{v}$, or in other words, an element lying in the normal plane to the curve (L) which passes through the given point. The general solution (78) also gives a family of surfaces whose tangent planes at every point are normal to $\mathbf{v}$. The surfaces (78) are thus orthogonal to the curves (L). If a family of curves (L) is given which fills space, a tangential vector $\mathbf{v}$ can be found at every point; if the length of the vector is taken as unity, its components are P, Q, R, and equation (77) can be obtained. Equation (81) now gives the condition for the given family of curves (L) to be orthogonal to some family of surfaces.

77. Change of variables in double integrals. We conclude the present section by deducing the formulae given in [57] for the change of variables in a double integral. Let the transformations be given:

$$x = \varphi(u, v); \quad y = \psi(u, v), \tag{82}$$

where we look on (x, y) and (u, v) as the Cartesian coordinates of points in a plane. Equations (82) give the transformation of the plane such that the point (u, v) becomes the point (x, y). Let (σ_1), (σ) be domains in the plane with contours (l_1), (l) respectively. We assume that : (1) the functions of (82) are continuous together with their first order derivatives in the domain (σ_1) up to (l_1); (2) equations (82) define a one-to-one correspondence between (σ_1) with contour (l_1) and (σ) with contour (l), i.e. for every point (u, v) of (σ_1) there is a corresponding single point (x, y) of (σ), and conversely, for every point of (σ) there is one corresponding point of (σ_1); (3) the Jacobian of the functions of (82) with respect to (u, v):

$$\frac{D(\varphi, \psi)}{D(u, v)} = \frac{\partial\varphi(u, v)}{\partial u} \cdot \frac{\partial\psi(u, v)}{\partial v} - \frac{\partial\varphi(u, v)}{\partial v} \cdot \frac{\partial\psi(u, v)}{\partial u} \tag{83}$$

preserves the same sign throughout (σ_1).

We shall say that the *correspondence between* (σ) *and* (σ_1) *is direct*, if, on making a counter-clockwise circuit round (l_1), the point (x, y) makes a counter-clockwise circuit round (l). The correspondence is called *reverse* in the opposite case, when a circuit round (l_1) implies a circuit in the opposite direction round (l). The area σ of the domain (σ) is given by the integral [68]:

$$\sigma = \int_{(l)} x \, \mathrm{d}y,$$

where the integration is carried out counter-clockwise.

We find on introducing the new variables in accordance with (82):

$$\sigma = \pm \int_{(l_1)} \varphi(u, v) \, \mathrm{d}\psi(u, v) = \pm \int_{(l_1)} \varphi \frac{\partial\psi}{\partial u} \mathrm{d}u + \varphi \frac{\partial\psi}{\partial v} \mathrm{d}v. \tag{84}$$

Let the integration round (l_1) be counter-clockwise. This direction is in fact obtained round (l_1) after the transformation if the correspondence is direct, so that we must take the $(+)$ sign in (84). If the correspondence is reverse, we get the opposite direction after transformation to that round (l); if we now write the $(-)$ sign in front, we can again integrate counter-clockwise.

We apply Green's formula (18) to (84), after setting $x = u$, $y = v$, $P = \varphi \, \partial\psi/\partial u$, $Q = \varphi \, \partial\psi/\partial v$. We have:

$$\frac{\partial Q}{\partial u} - \frac{\partial P}{\partial v} = \frac{D(\varphi, \psi)}{D(u, v)}, \tag{85}$$

and hence:

$$\sigma = \pm \iint_{(\sigma_1)} \frac{D(\varphi, \psi)}{D(u, v)} \mathrm{d}u \, \mathrm{d}v.$$

On applying the mean value theorem [61] to the double integral, we have:

$$\sigma = \pm\, \sigma_1 \left[\frac{D(\varphi, \psi)}{D(u, v)}\right]_{(u_0, v_0)}, \tag{86}$$

where we take the value of the Jacobian (83) at a point (u_0, v_0) of (σ_1). Since σ and σ_1 are positive, a consequence of this last expression is that, if the correspondence is direct, the Jacobian (83) is positive, whilst it is negative if the correspondence is reverse.

We now come to the formula for change of variables. Let $f(x, y)$ be a continuous function in (σ). Let (σ_1) be divided into sections $\tau_1', \tau_2', \ldots, \tau_n'$. By (82), there will be a corresponding division of (σ) into sections $\tau_1, \tau_2, \ldots, \tau_n$. We shall use the same symbols τ_k', τ_k to denote the areas of the sections. We have by (86):

$$\tau_k = \tau_k' \left| \frac{D(\varphi, \psi)}{D(u, v)} \right|_{(u_k, v_k)},$$

where (u_k, v_k) is a point of τ_k'. We have the corresponding point $x_k = \varphi(u_k, v_k)$, $y_k = \psi(u_k, v_k)$, and we can write:

$$\sum_{k=1}^{n} f(x_k, y_k)\, \tau_k = \sum_{k=1}^{n} f[\varphi(u_k, v_k), \psi(u_k, v_k)] \left| \frac{D(\varphi, \psi)}{D(u, v)} \right|_{(u_k, v_k)} \cdot \tau_k'.$$

On passing to the limit, we get the formula for change of variables in a double integral:

$$\iint\limits_{(\sigma)} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = \iint\limits_{(\sigma_1)} f[\varphi(u, v), \psi(u, v)] \left| \frac{D(\varphi, \psi)}{D(u, v)} \right| \mathrm{d}u\, \mathrm{d}v, \tag{87}$$

which is the same as (13) of [57].

We notice a consequence of (86). Let (σ_1) contract indefinitely to the point (u, v). The point (u_0, v_0) of (σ_1) now tends to (u, v), whilst (σ) contracts indefinitely to the point (x, y). We obtain from (86), on passing to the limit:

$$\left| \frac{D(\varphi, \psi)}{D(u, v)} \right| = \lim \frac{\sigma}{\sigma_1},$$

i.e. the limit of the ratio of areas is the absolute value of the Jacobian at the corresponding point, as we already mentioned in [57]. Similarly, if we look on the function of a single variable $x = f(u)$ as transforming points on a straight line, so that the point with abscissa u becomes a point with abscissa x, the absolute value of the derivative $|\, f'(u)\, |$ gives the limit of the ratio of corresponding lengths on the line,

or in other words, the coefficient of the linear change produced by the transformation at the point of abscissa u.

We remark that we had to use the second derivative $\partial^2\varphi/\partial u\, \partial v$ and its independence of the order of differentiation when obtaining (85). We must therefore, strictly speaking, add the further assumption to those at the beginning of the section of the existence and continuity of $\partial^2\varphi/\partial u\, \partial v$, whence follows its independence of the order of differentiation, as we know from [I, 155].

If (v) is a domain of space, bounded by the surface (S), we can apply Ostrogradskii's formula [63] with $P = Q = 0$ and $R = z$ to express the volume v of the domain as an integral over the surface:

$$v = \iint\limits_{(S)} z \cos(n, Z)\, dS.$$

By using this expression for the volume, we can prove the formula for change of variables in a triple integral [60] in much the same way as above for a double integral.

§ 8. Improper integrals and integrals that depend on a parameter

78. Integration under the integral sign. When evaluating multiple integrals we encountered definite integrals in which the integrands and even the limits of integration depend on a variable parameter. We now pause to consider these integrals in some detail.

We take the integral:

$$I(y) = \int\limits_{x_1}^{x_2} f(x, y)\, dx,$$

where x denotes the variable of integration and where the integrand also depends on the variable parameter y, on which the limits also depend. It is clear that in this case the result of integration $I(y)$ will in general be a function of y. Formula (7) of [56]:

$$\int\limits_{\alpha}^{\beta} I(y)\, dy = \int\limits_{\alpha}^{\beta} dy \int\limits_{x_1}^{x_2} f(x, y)\, dx = \int\limits_{a}^{b} dx \int\limits_{y_1}^{y_2} f(x, y)\, dy \tag{1}$$

is known as the *formula for integrating a definite integral with respect to the parameter under the integral sign*. It reduces to a very simple

form when the limits x_1 and x_2 are constants a, b, independent of y [56]:

$$\int_{\alpha}^{\beta} I(y)\,\mathrm{d}y = \int_{\alpha}^{\beta} \mathrm{d}y \int_{a}^{b} f(x,y)\,\mathrm{d}x = \int_{a}^{b} \mathrm{d}x \int_{\alpha}^{\beta} f(x,y)\,\mathrm{d}y. \tag{2}$$

The integrand $f(x, y)$ is assumed continuous with respect to both variables in the domain of integration in all these formulae, and the domain is assumed to be finite. We are concerned in other cases with improper multiple integrals, which are dealt with later.

Example. The above method is sometimes used to evaluate definite integrals when the indefinite integrals are unknown. We use it to evaluate Laplace's integral:

$$I = \int_0^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x. \tag{3}$$

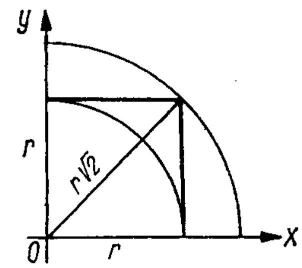

FIG. 71

Let (D') be the quadrant of a circle with radius r and centre at the origin lying in the first quadrant of the axes; let (D'') be the square bounded by the lines $x = 0$, $x = r$, $y = 0$, $y = r$; and finally, let (D''') be the quadrant of the circle with centre at the origin and radius $r\sqrt{2}$ (Fig. 71). Clearly, (D') is part of (D''), and (D'') is part of (D'''). We take the double integral over these domains of the positive function $\mathrm{e}^{-x^2-y^2}$, and have the obvious inequalities:

$$\iint_{(D')} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y < \iint_{(D'')} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y < \iint_{(D''')} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y.$$

We introduce polar coordinates $x = \varrho\cos\varphi$, $y = \varrho\sin\varphi$, and obtain [56]:

$$\iint_{(D')} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y = \int_0^{\frac{\pi}{2}} \mathrm{d}\varphi \int_0^{r} \mathrm{e}^{-\varrho^2}\varrho\,\mathrm{d}\varrho = \frac{\pi}{2}\left[-\frac{1}{2}\,\mathrm{e}^{-\varrho^2}\right]_{\varrho=0}^{\varrho=r} = \frac{\pi}{4}(1-\mathrm{e}^{-r^2}).$$

We have on replacing r by $r\sqrt{2}$:

$$\iint_{(D''')} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y = \frac{\pi}{4}(1-\mathrm{e}^{-2r^2}).$$

Integration over the square (D'') gives:

$$\iint_{(D'')} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y = \int_0^r \mathrm{e}^{-x^2}\,\mathrm{d}x \cdot \int_0^r \mathrm{e}^{-y^2}\,\mathrm{d}y = \left(\int_0^r \mathrm{e}^{-x^2}\,\mathrm{d}x\right)^2,$$

and the inequality written above takes the form:

$$\frac{\pi}{4}(1-\mathrm{e}^{-r^2}) < \left(\int \mathrm{e}^{-x^2}\,\mathrm{d}x\right)^2 < \frac{\pi}{4}(1-\mathrm{e}^{-2r^2}).$$

The two extreme terms of the inequality tend to $\pi/4$ as r tends to infinity, so that the middle term must tend to the same limit; hence we obtain for integral (3):

$$\int\limits_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \frac{\sqrt{\pi}}{2}. \tag{4}$$

It is easily seen [I, 96] that:

$$\int\limits_{-\infty}^{+\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x = 2\int\limits_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}. \tag{5}$$

If use is made of the improper integral over the whole of the first quadrant of the axes, denoted here by (P), the above result follows almost at once. We have:

$$\iint\limits_{(P)} \mathrm{e}^{-x^2-y^2}\mathrm{d}x\,\mathrm{d}y = \int\limits_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x \,.\, \int\limits_0^\infty \mathrm{e}^{-y^2}\,\mathrm{d}y = I^2,$$

and on introducing polar coordinates:

$$I^2 = \iint\limits_{(P)} \mathrm{e}^{-\varrho^2}\varrho\,\mathrm{d}\varrho\,\mathrm{d}\varphi = \int\limits_0^{\frac{\pi}{2}} \mathrm{d}\varphi \int\limits_0^\infty \mathrm{e}^{-\varrho^2}\varrho\,\mathrm{d}\varrho = \frac{\pi}{2}\left[-\frac{1}{2}\mathrm{e}^{-\varrho^2}\right]_{\varrho=0}^{\varrho=\infty} = \frac{\pi}{4},$$

whence $I = \sqrt{\pi}/2$, which agrees with the value found above.

79. Dirichlet's formula. If we specify x_1 and x_2 in (1) as functions of y and the interval (α, β) of variation of y, we thereby define a domain (σ) of the xy plane. A frequent case in applications is that when the domain becomes an isosceles triangle, formed by the three lines (Fig. 72):

$$y = x; \quad x = a; \quad y = b.$$

If we reduce a double integral over this triangle to an iterated integral, in one case by integrating first with respect to x, then with respect to y, and in the other case by integrating first with respect to y, then with respect to x, we arrive at:

$$\int\limits_a^b \mathrm{d}y \int\limits_a^y f(x, y)\,\mathrm{d}x = \int\limits_a^b \mathrm{d}x \int\limits_x^b f(x, y)\,\mathrm{d}y, \tag{6}$$

which is known as Dirichlet's formula.

Example. Abel's problem. *To find the curve in a vertical plane such that a material particle setting out with zero initial velocity from any point M of the curve*

of height h (Fig. 73) *and falling along the curve, reaches its lowest point O in a time T which is a given function of the height h:*

$$T = \varphi(h).$$

We take OY vertically upwards, OX horizontally, and the origin at the owest point of the curve; the equation of the curve is sought as:

$$x = f(y).$$

We write:

$$ds = dy\sqrt{1 + [f'(y)]^2} = u(y)\,dy; \qquad u(y) = \sqrt{1 + [f'(y)]^2}. \tag{7}$$

We know from dynamics that the increment in kinetic energy when the particle moves from the initial position M to N is equal to the work done by gravity,

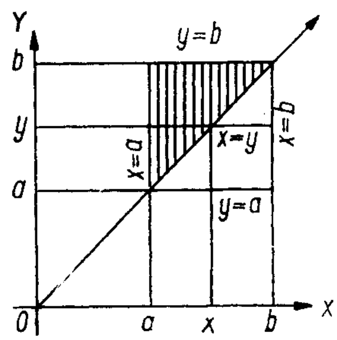

FIG. 72

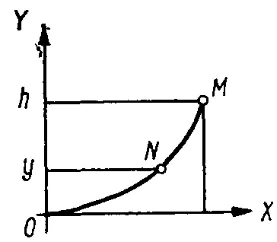

FIG. 73

since the reaction of the curve is perpendicular to the displacement of the particle and hence does no work; thus,

$$\frac{1}{2}mv^2 = mg(h - y); \quad v = \frac{ds}{dt},$$

or

$$\frac{1}{2}\left(\frac{ds}{dt}\right)^2 = g(h - y)$$

$$dt = \frac{-ds}{\sqrt{2g(h-y)}} = \frac{1}{\sqrt{2g}}\,\frac{-u(y)}{\sqrt{h-y}}\,dy,$$

where the $(-)$ sign is taken because the height y of the particle decreases with increasing t.

The time of descent from M to O corresponds to y varying from h to 0, so that:

$$\varphi(h) = T = \frac{1}{\sqrt{2g}}\int_0^h \frac{u(y)\,dy}{\sqrt{h-y}}. \tag{8}$$

We thus have to find $u(y)$ from equation (8); this is called an *integral equation* since the unknown $u(y)$ occurs under an integral sign.

We multiply both sides of (8) by $1/\sqrt{z-h}$, and integrate with respect to h between the limits 0 and z:

$$\int_0^z \frac{\varphi(h)}{\sqrt{z-h}}\,\mathrm{d}h = \frac{1}{\sqrt{2g}}\int_0^z \frac{\mathrm{d}h}{\sqrt{z-h}}\int_0^h \frac{u(y)\,\mathrm{d}y}{\sqrt{h-y}}\,.$$

The iterated integral on the right can be transformed as follows by Dirichlet's formula:

$$\begin{aligned}\int_0^z \frac{\mathrm{d}h}{\sqrt{z-h}}\int_0^h \frac{u(y)\,\mathrm{d}y}{\sqrt{h-y}} &= \int_0^z \mathrm{d}y \int_y^z \frac{u(y)}{\sqrt{(z-h)(h-y)}}\,\mathrm{d}h = \\ &= \int_0^z u(y)\,\mathrm{d}y \int_y^z \frac{\mathrm{d}h}{\sqrt{(z-h)(h-y)}}\,.\end{aligned} \tag{9}$$

There is no great difficulty in evaluating the inner integral, if a new variable t is introduced in accordance with:

$$h = y + t(z-y).$$

When h varies from y to z, t varies from 0 to 1, and we have:

$$z-h = (z-y)(1-t);\quad h-y = (z-y)t;\quad \mathrm{d}h = (z-y)\,\mathrm{d}t,$$

whence

$$\int_y^z \frac{\mathrm{d}h}{\sqrt{(z-h)(h-y)}} = \int_0^1 \frac{\mathrm{d}t}{\sqrt{t(1-t)}} = \int_0^1 \frac{\mathrm{d}t}{\sqrt{\frac{1}{4}-\left(t-\frac{1}{2}\right)^2}} =$$

$$= \text{arc sin}(2t-1)\Big|_{t=0}^{t=1} = \text{arc sin } 1 - \text{arc sin }(-1) = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi$$

and finally:

$$\frac{\pi}{\sqrt{2g}}\int_0^z u(y)\,\mathrm{d}y = \int_0^z \frac{\varphi(h)\,\mathrm{d}h}{\sqrt{z-h}}\,,$$

or

$$\int_0^z u(y)\,\mathrm{d}y = F(z), \tag{10}$$

where $F(z)$ is a known function of z, given by:

$$F(z) = \frac{\sqrt{2g}}{\pi}\int_0^z \frac{\varphi(h)\,\mathrm{d}h}{\sqrt{z-h}}\,.$$

We find on differentiating (10) with respect to z:

$$u(z) = \frac{\mathrm{d}F(z)}{\mathrm{d}z} = \frac{\sqrt{2g}}{\pi} \frac{\mathrm{d}}{\mathrm{d}z} \int_0^z \frac{\varphi(h)\,\mathrm{d}h}{\sqrt{z-h}}, \tag{11}$$

which solves the problem, since, on knowing $u(y)$, $x = f(y)$ may be found without difficulty from (7).

We work out the problem in full for the particular case of the *isochronous curve*, for which the general time of descent is independent of the height h, i.e.

$$\varphi(h) = \text{const.} = c.$$

We have here:

$$F(z) = \frac{\sqrt{2g}}{\pi} \int_0^z \frac{c\,\mathrm{d}h}{\sqrt{z-h}} = \frac{c\sqrt{2g}}{\pi} 2\sqrt{z},$$

$$u(z) = \frac{c\sqrt{2g}}{\pi\sqrt{z}}.$$

From (7), we have the following equation from which $x = f(y)$ can be determined:

$$(\mathrm{d}x)^2 + (\mathrm{d}y)^2 = \frac{2gc^2}{\pi^2} \frac{(\mathrm{d}y)^2}{y} = \frac{A}{y} (\mathrm{d}y)^2 \quad \left(A = \frac{2gc^2}{\pi^2}\right).$$

We write:

$$y = a(1 + \cos t); \quad \mathrm{d}y = -a \sin t\,\mathrm{d}t; \quad A = 2a.$$

We now find:

$$\mathrm{d}x = \mathrm{d}y \sqrt{\frac{2a}{y} - 1} = \sqrt{\frac{1-\cos t}{1+\cos t}} \cdot (-a \sin t)\,\mathrm{d}t = -2a \sin^2 \frac{t}{2}\,\mathrm{d}t;$$

$$x = x_0 - a(t - \sin t),$$

where x_0 is a constant of integration. The reader can easily show that the curve is a cycloid, in a slightly different disposition to the cycloid of [I, 79].

We show later how the differentiation with respect to z is carried out in the general formula (11).

We notice some points regarding the solution obtained. The integral equation (8) was solved on the basis of the assumption that a solution exists. Solution (11) ought to be checked, strictly speaking, i.e. (11) substituted for $u(z)$ in (8) and the equality of the left and right-hand sides shown. A further point: double integral (9) is improper in the sense that the integrand tends to infinity. We shall see later that (9) exists, and we can easily show that expression (1) for reducing it to an iterated integral, is applicable in this case.

80. Differentiation under the integral sign. We take an integral depending on a parameter y:

$$I(y) = \int_a^b f(x, y)\,\mathrm{d}x. \tag{12}$$

We take the limits a and b as at present independent of y. We assume that $f(x, y)$ is continuous and has a continuous partial derivative $\partial f(x, y)/\partial y$ in the rectangle: $a \leqslant x \leqslant b$; $\alpha \leqslant y \leqslant \beta$. We show that with these assumptions a derivative $\mathrm{d}I(y)/\mathrm{d}y$ exists which can be obtained by *differentiation with respect to y under the integral sign*, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}y}\int_a^b f(x, y)\,\mathrm{d}x = \int_a^b \frac{\partial f(x, y)}{\partial y}\,\mathrm{d}x. \tag{13}$$

The increment $\Delta I(y)$ of $I(y)$ is given by:

$$\Delta I(y) = I(y + \Delta y) - I(y) = \int_a^b [f(x, y + \Delta y) - f(x, y)]\,\mathrm{d}x. \tag{14}$$

We get from the formula for finite increments:

$$f(x, y + \Delta y) - f(x, y) = \Delta y\,\frac{\partial f(x, y + \theta\,\Delta y)}{\partial y}\,(0 < \theta < 1). \tag{15}$$

We use the uniform continuity of $\partial f(x, y)/\partial y$ in the rectangle mentioned to write

$$\frac{\partial f(x, y + \theta\Delta y)}{\partial y} = \frac{\partial f(x, y)}{\partial y} + \eta(x, y, \Delta y) \tag{16}$$

where $\eta(x, y, \Delta y)$ tends uniformly to zero with respect to x and y as $\Delta y \to 0$, i.e. given any positive ε, there exists a δ such that $|\eta(x, y, \Delta y)| < \varepsilon$ provided $|\Delta y| < \delta$. Hence it follows that

$$\left|\int_a^b \eta(x, y, \Delta y)\,\mathrm{d}x\right| \leqslant \int_a^b \varepsilon\,\mathrm{d}x = \varepsilon(b - a)\ (|\Delta y| < \delta),$$

and since ε is arbitrarily small, we have

$$\int_a^b \eta(x, y, \Delta y)\,\mathrm{d}x \to 0 \text{ as } \Delta y \to 0. \tag{17}$$

We return to (14). Noticing that Δy is independent of x, and using (15) and (16), we can write:

$$\Delta I(y) = \Delta y \int_a^b \frac{\partial f(x, y)}{\partial y}\,\mathrm{d}x + \Delta y \int_a^b \eta(x, y, \Delta y)\,\mathrm{d}x.$$

On dividing by Δy and passing to the limit, we get by (17):

$$\lim_{\Delta y \to 0} \frac{\Delta I(y)}{\Delta y} = \int_a^b \frac{\partial f(x, y)}{\partial y}\,\mathrm{d}x,$$

i.e. we have proved (13). We remark that if the continuity of $f(x, y)$ only is assumed, it still follows, from (14) and the fact that $[f(x, y + \Delta y) - f(x, y)]$ tends uniformly to zero with respect to x and y as $\Delta y \to 0$, that $I(y)$ is a continuous function of y.

With the same assumptions regarding $f(x, y)$, we now consider

$$I_1(y) = \int_{x_1}^{x_2} f(x, y)\,\mathrm{d}x, \tag{18}$$

where the limits of integration belong to the interval (a, b) and are assumed continuous functions of y, with derivatives with respect to y.

Let Δx_1, Δx_2 be the respective increments of x_1, x_2 when y receives the increment Δy.

We have:

$$\Delta I_1(y) = I_1(y + \Delta y) - I_1(y) =$$
$$= \int_{x_1+\Delta x_1}^{x_2+\Delta x_2} f(x, y + \Delta y)\,\mathrm{d}x - \int_{x_1}^{x_2} f(x, y)\,\mathrm{d}x. \tag{19}$$

On recalling [I, 94] that:

$$\int_{x_1+\Delta x_1}^{x_2+\Delta x_2} = \int_{x_1}^{x_2} + \int_{x_2}^{x_2+\Delta x_2} - \int_{x_1}^{x_1+\Delta x_1},$$

we can re-arrange (19) as:

$$\Delta I_1(y) = \int_{x_1}^{x_2} [f(x, y + \Delta y) - f(x, y)]\,\mathrm{d}x +$$
$$+ \int_{x_2}^{x_2+\Delta x_2} f(x, y + \Delta y)\,\mathrm{d}x - \int_{x_1}^{x_1+\Delta x_1} f(x, y + \Delta y)\,\mathrm{d}x. \tag{20}$$

We naturally assume in these arguments that $f(x, y)$ satisfies the conditions mentioned above for $\alpha \leqslant y \leqslant \beta$ and for all x belonging to the intervals of integration.

We can write, by the mean value theorem [I, 95]:

$$\int_{x_1}^{x_1+\Delta x_1} f(x, y + \Delta y)\,\mathrm{d}x = \Delta x_1\, f(x_1 + \theta_1\,\Delta x_1, y + \Delta y) =$$
$$= \Delta x_1\,[f(x_1, y) + \eta_1]$$
$$\int_{x_2}^{x_2+\Delta x_2} f(x, y + \Delta y)\,\mathrm{d}x = \Delta x_2\, f(x_2 + \theta_2\,\Delta x_2, y + \Delta y) =$$
$$= \Delta x_2\,[f(x_2, y) + \eta_2]$$
$$(0 < \theta_1 \text{ and } \theta_2 < 1).$$

If $\Delta y \to 0$, Δx_1, $\Delta x_2 \to 0$, and we can say, by the continuity of $f(x, y)$, that now η_1, $\eta_2 \to 0$.

On substituting these expressions in (20) and using (15) and (16), we obtain after division by Δy:

$$\frac{\Delta I_1(y)}{\Delta y} = \int_{x_1}^{x_2} \frac{\partial f(x, y)}{\partial y} \mathrm{d}x + [f(x_2, y) + \eta_2] \frac{\Delta x_2}{\Delta y} -$$

$$- [f(x_1, y) + \eta_1] \frac{\Delta x_1}{\Delta y} + \int_{x_1}^{x_2} \eta(x, y, \Delta y) \mathrm{d}x.$$

Passage to the limit gives us, by (17), the following formula for differentiating integral (18):

$$\frac{\mathrm{d}}{\mathrm{d}y} \int_{x_1}^{x_2} f(x, y) \mathrm{d}x = \int_{x_1}^{x_2} \frac{\partial f(x, y)}{\partial y} \mathrm{d}y + f(x_2, y) \frac{\mathrm{d}x_2}{\mathrm{d}y} - f(x_1, y) \frac{\mathrm{d}x_1}{\mathrm{d}y}. \quad (21)$$

This becomes (13) if x_1 and x_2 are independent of y. Formula (13) is also valid for differentiation of a multiple integral with respect to a parameter, provided the domain of integration (B) is independent of the parameter. If, for instance, the integrand $f(M, t)$ in a double integral over (B) depends on the parameter t as well as on the variable point M, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \iint_{(B)} f(M, t) \mathrm{d}\sigma = \iint_{(B)} \frac{\partial f(M, t)}{\partial t} \mathrm{d}\sigma. \quad (22)$$

It is assumed here that $f(M, t)$ and $\partial f(M, t)/\partial t$ are continuous for M varying in (B) including its contour and for t varying in a given interval.

The interval of integration must be finite in the proofs of (13) and (22). We apply (13) for an infinite interval in the examples, whilst indicating later the conditions for the validity of the procedure.

It also follows from the above formulae that, if $f(x, y)$, $x_2(y)$ and $x_1(y)$ are continuous, integral (18) is likewise a continuous function of y.

81. Examples. 1. We found in [28] the particular solution of the equation

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t} + k^2 y = f(t),$$

satisfying the conditions

$$y\bigg|_{t=0} = \frac{\mathrm{d}y}{\mathrm{d}t}\bigg|_{t=0} = 0. \quad (23)$$

It has the form:

$$y = \frac{1}{k}\int_0^t f(u)\sin k(t-u)\,\mathrm{d}u.$$

This may easily be checked by direct differentiation in accordance with rule (21). We have:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \int_0^t f(u)\cos k(t-u)\,\mathrm{d}u + \frac{1}{k} f(u)\sin k(t-u)\Big|_{u=t} = \int_0^t f(u)\cos k(t-u)\,\mathrm{d}u$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = -k\int_0^t f(u)\sin k(t-u)\,\mathrm{d}u + f(u)\cos k(t-u)\Big|_{u=t} = -k^2y + f(t).$$

i.e. in fact:

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2} + k^2y = f(t).$$

The expressions above give conditions (23) directly, on setting $t = 0$.

2. Let us evaluate the integral [I, 110]:

$$I_1 = \int\frac{\log(1+x)}{1+x^2}\,\mathrm{d}x.$$

We introduce a parameter a and consider:

$$I(a) = \int_0^a \frac{\log(1+ax)}{1+x^2}\,\mathrm{d}x.$$

It is immediately clear that

$$I(0) = 0 \quad\text{and}\quad I(1) = I_1.$$

Application of (21) in regard to a gives:

$$\frac{\mathrm{d}I(a)}{\mathrm{d}a} = \int_0^a \frac{x}{(1+ax)(1+x^2)}\,\mathrm{d}x + \frac{\log(1+a^2)}{1+a^2}.$$

We get on decomposing the rational fraction into partial fractions:

$$\frac{x}{(1+ax)(1+x^2)} = \frac{1}{1+a^2}\left[-\frac{a}{1+ax} + \frac{x}{1+x^2} + \frac{a}{1+x^2}\right],$$

and on integrating with respect to x:

$$\int_0^a \frac{x}{(1+ax)(1+x^2)}\,\mathrm{d}x = -\frac{\log(1+a^2)}{2(1+a^2)} + \frac{a\arctan a}{1+a^2}.$$

Finally:

$$\frac{\mathrm{d}I(a)}{\mathrm{d}a} = -\frac{\log(1+a^2)}{2(1+a^2)} + \frac{a\arctan a}{1+a^2} + \frac{\log(1+a^2)}{1+a^2} = \frac{\log(1+a^2)}{2(1+a^2)} + \frac{a\arctan a}{1+a^2}.$$

$$I(a) = \frac{1}{2}\int\frac{\log(1+a^2)}{1+a^2}\,\mathrm{d}a + \int\frac{a\arctan a}{1+a^2}\,\mathrm{d}a, \tag{24}$$

where we do not write a constant of integration since $I(0) = 0$. The second term may be integrated by parts:

$$\int_0^a \frac{\alpha \arctan \alpha}{1+\alpha^2}\, d\alpha = \frac{1}{2}\int_0^a \arctan \alpha \, d \log (1+\alpha^2) =$$

$$= \frac{1}{2} \arctan \alpha \cdot \log (1+\alpha^2) \Big|_{\alpha=0}^{\alpha=a} - \frac{1}{2}\int_0^a \frac{\log (1+\alpha^2)}{1+\alpha^2}\, d\alpha,$$

and therefore, by (24):

$$I(a) = \frac{1}{2} \arctan a . \log (1+a^2),$$

whence, with $a = 1$:

$$I_1 = \int_0^1 \frac{\log(1+x)}{1+x^2}\, dx = \frac{\pi}{8} \log 2.$$

3. We evaluate the integral:

$$\int_0^\infty \frac{\sin \beta x}{x}\, dx.$$

We consider instead of this what at first sight appears to be more difficult:

$$I(\alpha, \beta) = \int_0^\infty e^{-\alpha x} \frac{\sin \beta x}{x}\, dx \qquad (\alpha > 0). \tag{25}$$

We differentiate with respect to β:

$$\frac{\partial I(\alpha,\beta)}{\partial \beta} = \int_0^\infty \frac{\partial}{\partial \beta}\left(e^{-\alpha x} \frac{\sin \beta x}{x}\right) dx = \int_0^\infty e^{-\alpha x} \cos \beta x \, dx,$$

This last integral is easily evaluated [I, 201]:

$$\frac{\partial I(\alpha,\beta)}{\partial \beta} = \int_0^\infty e^{-\alpha x} \cos \beta x\, dx = e^{-\alpha x} \frac{-\alpha \cos \beta x + \beta \sin \beta x}{\alpha^2 + \beta^2} \Big|_{x=0}^{x=\infty} = \frac{\alpha}{\alpha^2+\beta^2},$$

whence

$$I(\alpha,\beta) = \int \frac{\alpha\, d\beta}{\alpha^2+\beta^2} + C = \arctan \frac{\beta}{\alpha} + C, \tag{26}$$

It only remains to find the constant of integration C, which is independent of β. We let β tend to zero in (25) and (26):

$$\lim_{\beta \to 0} I(\alpha,\beta) = I(\alpha, 0) = 0; \quad I(\alpha, 0) = \arctan 0 + C = 0,$$

whence it is clear that $C = 0$. We thus have:

$$I(\alpha,\beta) = \arctan \frac{\beta}{\alpha}.$$

The integral to be evaluated is obtained from $I(a, \beta)$ with $a = 0$, where a must tend to zero through positive values, i.e. $a \to +0$. If we let a approach zero in the above equation, we get different limits according to whether $\beta > 0$ or $\beta < 0$:

$$\lim_{a \to +0} \operatorname{arc\,tan} \frac{\beta}{a} = \begin{cases} \frac{1}{2}\pi & \text{for } \beta > 0 \\ -\frac{1}{2}\pi & \text{for } \beta < 0 \\ 0 & \text{for } \beta = 0 \end{cases}$$

so that finally:†

$$I(\beta) = \int_0^\infty \frac{\sin \beta x}{x}\,dx = \begin{cases} \frac{1}{2}\pi & \text{for } \beta > 0 \\ -\frac{1}{2}\pi & \text{for } \beta < 0 \\ 0 & \text{for } \beta = 0. \end{cases}$$

The integral gives us a discontinuous function $I(\beta)$ of β. The graph of the function consists of two half-lines and a point, and is shown in Fig. 74.

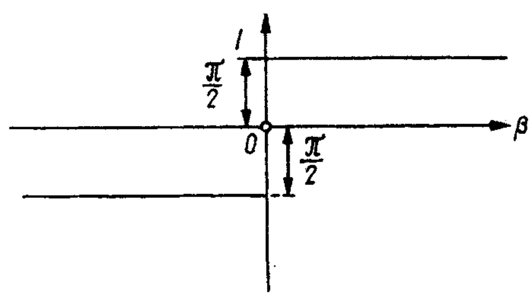

FIG. 74

4. On differentiating the obvious equation

$$\int_0^\infty e^{-ax}dx = \frac{1}{a} \qquad (a > 0)$$

† The above arguments are not rigorous, since the equations have been assumed: $\lim_{\beta \to +0} I(a, \beta) = I(a, 0)$; $\lim_{a \to 0} I(a, \beta) = = I(0, \beta)$; these can be taken as self-evident if $I(a, \beta)$ is known to be continuous with respect to both β and a. A further point is that, if we had not introduced the factor e^{-ax} under the integral sign, the meaningless integral $\int_0^\infty \cos \beta x\,dx$ would have been obtained after differentiation. A rigorous proof of the continuity of $I(a, \beta)$ is given in [85].

k times with respect to a, we get:

$$\int_0^\infty e^{-ax} x^k \, dx = \frac{k!}{a^{k+1}} .$$

We now consider the integral:

$$I_n = \int_0^\infty e^{-ax^2} x^n \, dx \quad (a > 0).$$

If n is odd: $n = 2k + 1$, I_n is obtained by substituting $x^2 = t$:

$$I_{2k+1} = \int_0^\infty e^{-ax^2} x^{2k} x \, dx = \frac{1}{2} \int_0^\infty e^{-at} t^k \, dt = \frac{1}{2} \frac{k!}{a^{k+1}} .$$

For the case of n even, we introduce a new variable of integration $x = \sqrt{at}$ into (4). After replacing t by x again in the result obtained, we arrive at:

$$I_0 = \int_0^\infty e^{-ax^2} \, dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} .$$

We find on differentiating this k times with respect to a:

$$\frac{d^k I_0}{da^k} = (-1)^k \int_0^\infty e^{-ax^2} x^{2k} \, dx,$$

whence

$$I_{2k} = (-1)^k \frac{d^k}{da^k} \left(\frac{1}{2} \sqrt{\frac{\pi}{a}} \right) = \frac{\sqrt{\pi}}{2} \cdot \frac{1 \cdot 3 \ldots (2k-1)}{2^k \cdot a^{k+\frac{1}{2}}} .$$

5. We shall take a as constant in the integral:

$$I(\beta) = \int_0^\infty e^{-ax^2} \cos \beta x \, dx \qquad (a > 0),$$

which depends on two parameters a and β. We differentiate with respect to β:

$$\frac{dI(\beta)}{d\beta} = -\int_0^\infty e^{-ax^2} \sin \beta x \cdot x \, dx = \frac{1}{2a} \int_0^\infty \sin \beta x \, de^{-ax^2} .$$

We now integrate by parts:

$$\frac{dI(\beta)}{d\beta} = \frac{1}{2a} e^{-ax^2} \sin \beta x \Big|_{x=0}^{x=\infty} - \frac{\beta}{2a} \int_0^\infty e^{-ax^2} \cos \beta x \, dx = -\frac{\beta}{2a} \int_0^\infty e^{-ax_2} \cos \beta x \, dx,$$

that is,

$$\frac{dI(\beta)}{d\beta} = -\frac{\beta}{2a} I(\beta) .$$

This is a differential equation with separable variables:

$$\frac{dI(\beta)}{I(\beta)} = -\frac{\beta}{2a}\,d\beta,$$

whence we obtain on integration:

$$I(\beta) = Ce^{-\frac{\beta^2}{4a}}, \tag{27}$$

where the constant C is independent of β. We substitute $\beta = 0$:

$$I(0) = \int_0^\infty e^{-ax^2}\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}.$$

On the other hand, by (27):

$$I(0) = C,$$

so that $C = 1/2\sqrt{\pi/a}$; and finally, on substituting this value for C in (27):

$$\int_0^\infty e^{-ax^2}\cos\beta x\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\,e^{-\frac{\beta^2}{4a}}$$

On replacing a by a^2, we get a result that will be used later in investigating the equation of thermal radiation:

$$\int_0^\infty e^{-a_2x^2}\cos\beta x\,dx = \frac{\sqrt{\pi}}{2a}\,e^{-\frac{\beta^2}{4a^2}}.$$

82. Improper integrals. We have come across integrals more than once in which either the integrands or the limits have tended to infinity. We agreed in [I, 97,98] to ascribe a definite meaning to such integrals, provided certain conditions were fulfilled. We now treat the subject in detail.

1. The integrand tends to infinity. In the integral:

$$\int_a^b f(x)\,dx \qquad (b > a)$$

let $f(x)$ be continuous for $a \leqslant x < b$, whilst tending to infinity at $x = b$; or more precisely, let $f(x)$ be unbounded as x tends to b from below. We now have by definition [I, 97]:

$$\int_a^b f(x)\,dx = \lim_{\varepsilon\to+0}\int_a^{b-\varepsilon} f(x)\,dx,$$

provided the limit on the right exists. The condition for existence may be investigated as follows. The necessary and sufficient con-

dition for a variable to have a limit is, by Cauchy's basic test [I, 31], that the difference between any two values of the variable, as from a certain point of its variation, should be less in absolute value than any previously assigned positive number. The difference becomes in our case:

$$\int_a^{b-\varepsilon''} f(x)\,\mathrm{d}x - \int_a^{b-\varepsilon'} f(x)\,\mathrm{d}x = \int_{b-\varepsilon'}^{b-\varepsilon''} f(x)\,\mathrm{d}x \qquad (\varepsilon'' < \varepsilon'),$$

and we arrive at the following general condition: *a necessary and sufficient condition for the existence (convergence) of the improper integral*

$$\int_a^b f(x)\,\mathrm{d}x,$$

in which the integrand $f(x)$ tends to infinity for $x = b - 0$, is that for any given small positive δ there exists an η such that

$$\left|\int_{b-\varepsilon'}^{b-\varepsilon''} f(x)\,\mathrm{d}x\right| < \delta \quad \text{for} \quad 0 < \varepsilon' \quad \text{and} \quad \varepsilon'' < \eta.$$

If we use the familiar inequality [I, 95]:

$$\left|\int_{b-\varepsilon'}^{b-\varepsilon''} f(x)\,\mathrm{d}x\right| \leqslant \int_{b-\varepsilon'}^{b-\varepsilon''} |f(x)|\,\mathrm{d}x,$$

it follows immediately that the convergence of the integral

$$\int_a^b |f(x)|\,\mathrm{d}x \tag{28}$$

implies the convergence of

$$\int_a^b f(x)\,\mathrm{d}x. \tag{29}$$

The converse is not true, i.e. the convergence of integral (29) does not imply the convergence of (28). If (28) converges, (29) is said to be absolutely convergent [cf. I, 124].

A test of Cauchy's which is very important in applications follows from the general test: *if the integrand $f(x)$ is continuous for $a \leqslant x < b$ and, for x near b, satisfies the condition*

$$|f(x)| < \frac{A}{(b-x)^p}, \tag{30}$$

where A and p are positive constants and $p < 1$, the improper integral (29) *is absolutely convergent.*

On the other hand, if

$$|f(x)| > \frac{A}{(b-x)^p} \quad \text{and} \quad p \geqslant 1, \tag{31}$$

integral (29) *does not exist.*

We have, in fact, in case (30):

$$\left|\int_{b-\varepsilon'}^{b-\varepsilon''} f(x)\,\mathrm{d}x\right| \leqslant \int_{b-\varepsilon'}^{b-\varepsilon''} |f(x)|\,\mathrm{d}x < A\int_{b-\varepsilon'}^{b-\varepsilon''} \frac{\mathrm{d}x}{(b-x)^p} = A\,\frac{\varepsilon'^{1-p} - \varepsilon''^{1-p}}{1-p},$$

where the right-hand side becomes as small as desired, with sufficiently small ε' and ε'', since the power $(1-p)$ is positive $(p < 1)$.

In case (31), we can say first of all that the continuous function $f(x)$ keeps a constant sign in the neighbourhood of $x = b$; this follows because the absolute value of $f(x)$ remains greater than a positive number, by (31), so that $f(x)$ cannot vanish, as would be the case if it changed sign. If we confine ourselves to the case of positive $f(x)$, we have

$$\int_{b-\varepsilon'}^{b-\varepsilon''} f(x)\,\mathrm{d}x > A\int_{b-\varepsilon'}^{b-\varepsilon''} \frac{\mathrm{d}x}{(b-x)^p} =$$

$$= \begin{cases} A\log\dfrac{\varepsilon'}{\varepsilon''} & \text{for } p = 1 \\ A\,\dfrac{\varepsilon'^{(1-p)} - \varepsilon''^{(1-p)}}{1-p}, \end{cases}$$

where the right-hand side can be made as large as desired with suitably small ε' and ε'', since by hypothesis, $1 - p < 0$.

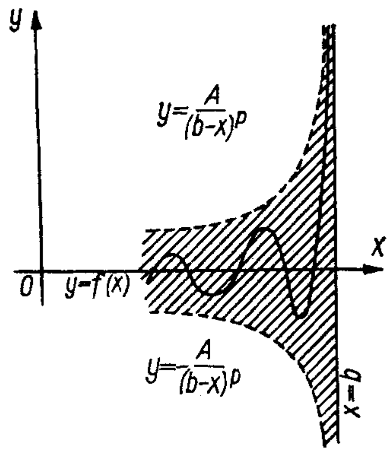

FIG. 75

The geometrical interpretation of Cauchy's test is very simple. In case (30), as x approaches b, the curve $y = f(x)$ lies wholly inside the area included between the two symmetrical curves

$$y = \pm\frac{A}{(b-x)^p} \tag{32}$$

(Fig. 75); this area is finite for $p < 1$, so that the area under $f(x)$ is also finite. In case (31), the curve $y = f(x)$ passes outside the shaded

area, and since this is infinite for curves (32) with $p \geqslant 1$, the area under $f(x)$ must also be infinite (Fig. 76).

Exactly similar arguments apply when $f(x)$ tends to infinity at the lower limit $x = a$, or at some intermediate point $x = c$ of the interval of integration [I, 97].

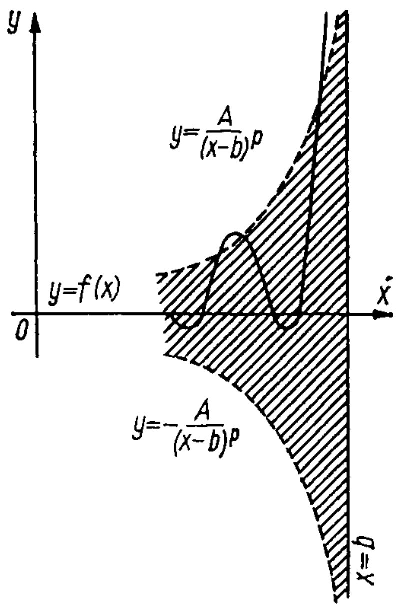

Fig. 76

2. Infinite limits. We now take the case of $b = +\infty$, i.e. the improper integral

$$\int_a^{+\infty} f(x)\,dx = \lim_{b \to +\infty} \int_a^b f(x)\,dx$$

with the assumption that $f(x)$ is continuous for $x \geqslant a$. We get on applying Cauchy's test as in the previous case: *the necessary and sufficient condition for the existence (convergence) of the improper integral*

$$\int_a^{+\infty} f(x)\,dx \tag{33}$$

is that, for any given small positive δ there exists a positive N such that

$$\left|\int_{b'}^{b''} f(x)\,dx\right| < \delta \text{ for } b' \text{ and } b'' > N.$$

In particular, we can prove Cauchy's test just as in the previous case: *if the integrand $f(x)$ is continuous for $x \geqslant a$ and*

$$|f(x)| < \frac{A}{x^p} \text{ and } p > 1, \tag{34}$$

the improper integral (33) *is absolutely convergent.*

If, however,

$$|f(x)| > \frac{A}{x^p} \text{ and } p \leqslant 1, \tag{35}$$

integral (33) *does not exist.*

The treatment is similar for the improper integrals [I, 98]:

$$\int_{-\infty}^{b} f(x)\,dx \text{ and } \int_{-\infty}^{+\infty} f(x)\,dx.$$

We indicate a convenient practical method for applying Cauchy's test. We start with an integral of type (33). Condition (34) for its convergence amounts to the existence of a $p > 1$ such that $f(x)\,x^p$ remains bounded as $x \to +\infty$. This is evidently true if a finite limit exists:

$$\lim_{x \to +\infty} f(x)\,x^p.$$

Similarly, condition (35) for divergence is satisfied if the limit exists:

$$\lim_{x \to +\infty} f(x)\, x^p \qquad (p \leqslant 1),$$

differing from zero (finite or infinite). Thus, e.g., the integral of example 5 [81] is absolutely convergent, since $e^{-\alpha x^2} \cos \beta x \cdot x^p$ tends to zero as $x \to +\infty$ for any positive p. This follows because the absolute value of $\cos \beta x$ does not exceed unity, whilst it is easily shown that $e^{-\alpha x^2} x^p \to 0$ by applying l'Hôpital's rule with, say, $p = 2$ [I, 65].

The integral:

$$\int_0^\infty \frac{5x^2 + 1}{x^3 + 4}\, dx$$

is divergent, since

$$\lim_{x \to +\infty} \frac{5x^2 + 1}{x^3 + 4}\, x = 5 \qquad (p = 1).$$

The integral of a rational fraction, with one or both limits infinite, is in general convergent only when the difference in the degrees of denominator and numerator is at least two. Furthermore, the denominator must not vanish in the interval of integration after any possible cancelling. If the interval is $(-\infty, +\infty)$, the denominator must have no real zeros.

Similar use can be made of conditions (30) and (31) for the convergence and divergence of integrals in which the integrand tends to infinity. For instance,

$$\int_0^1 \frac{\sin x}{x^m}\, dx$$

is convergent for $m < 2$, since $(\sin x/x^m)x^{m-1} = \sin x/x$ tends to unity as $x \to +0$, and $p = m - 1 < 1$. On the contrary, the integral is divergent for $m \geqslant 2$.

83. Conditionally convergent integrals. Cauchy's test only gives sufficient conditions (30) or (34) for the convergence of improper integrals. It cannot be applied say to conditionally convergent integrals, i.e. those where

$$\int_a^b f(x)\, dx \quad \text{or} \quad \int_a^{+\infty} f(x)\, dx$$

is convergent, whilst

$$\int_a^b |f(x)|\, dx \quad \text{or} \quad \int_a^{\infty} |f(x)|\, dx$$

is not convergent. We now give a test suitable for conditionally convergent integrals: *if*

$$F(x) = \int_a^x f(t)\, dt \qquad (a > 0)$$

remains bounded on indefinite increase of x,

$$\int_a^{+\infty} \frac{f(x)}{x^p}\,\mathrm{d}x$$

is convergent for any $p > 0$. We have, on integrating by parts:

$$\int_a^N \frac{f(x)}{x^p}\,\mathrm{d}x = \int_a^N \frac{1}{x^p}\,\mathrm{d}F(x) = \frac{F(x)}{x^p}\Big|_{x=a}^{x=N} + p\int_a^N \frac{F(x)}{x^{1+p}}\,\mathrm{d}x$$

or, since $F(a) = 0$:

$$\int_a^N \frac{f(x)}{x^p}\,\mathrm{d}x = \frac{F(N)}{N^p} + p\int_a^N \frac{F(x)}{x^{1+p}}\,\mathrm{d}x.$$

The first term on the right tends to zero with indefinite increase of N, since $F(N)$ is bounded by hypothesis and $p > 0$. The second term represents an integral which is convergent by Cauchy's test, since the numerator $F(x)$ of the integrand is bounded by hypothesis as $x \to +\infty$, whilst the degree of x in the denominator is greater than unity. Hence the limit exists:

$$\int_a^\infty \frac{f(x)}{x^p}\,\mathrm{d}x = \lim_{N\to\infty}\int_a^N \frac{f(x)}{x^p}\,\mathrm{d}x = p\int_a^{+\infty} \frac{F(x)}{x^{1+p}}\,\mathrm{d}x.$$

Examples. **1.** We take

$$\int_0^\infty \frac{\sin\beta x}{x}\,\mathrm{d}x, \tag{36}$$

already investigated in Example 3 [81]. We notice that the integrand takes a finite value β for $x = 0$, so that the integral is only improper on account of the infinite limit. It is clear that

$$\int_a^N \sin\beta x\,\mathrm{d}x = \left[-\frac{1}{\beta}\cos\beta x\right]_{x=a}^{x=N},$$

whence

$$\left|\int_a^N \sin\beta x\,\mathrm{d}x\right| \leqslant \frac{2}{\beta} \qquad (\beta > 0),$$

i.e. $\int_a^N \sin\beta x\,\mathrm{d}x$ is bounded for any a and N. The theorem proved is therefore applicable to (36), and this is convergent.

2. We also consider

$$\int_0^{\infty} \sin (x^2)\, dx. \tag{37}$$

On changing the variables in accordance with $x = \sqrt{t}$, this is reduced to the form

$$\frac{1}{2}\int_0^{\infty} \frac{\sin t}{\sqrt{t}}\, dt$$

and we can show its convergence as in Example 1. We consider in further detail the reasons for the convergence of (37). The graph of the integrand $f(x) = \sin (x^2)$

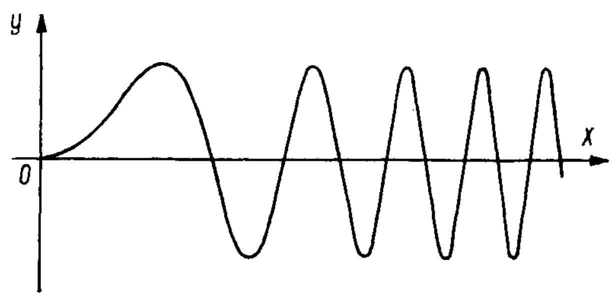

FIG. 77

is shown in Fig. 77; since $f(x)$ does not tend to zero as $x \to +\infty$, Cauchy's test cannot be applied. We subdivide the interval $(0, +\infty)$ as follows:

$$(0, \sqrt{\pi});\ (\sqrt{\pi}, \sqrt{2\pi});\ (\sqrt{2\pi}, \sqrt{3\pi});\ \ldots;\ (\sqrt{n\pi}, \sqrt{(n+1)\pi});\ \ldots,$$

so that $f(x)$ retains an invariable sign in each sub-interval: $(+)$ in the first, $(-)$ in the second, $(+)$ in the third, etc. We write:

$$u_n = (-1)^n \int_{\sqrt{n\pi}}^{\sqrt{(n+1)\pi}} \sin (x^2)\, dx.$$

On replacing x by a new variable t:

$$x = \sqrt{t + n\pi},$$

we get:

$$u_n = \frac{(-1)^n}{2}\int_0^{\pi} \frac{\sin (t + n\pi)}{\sqrt{t + n\pi}}\, dt = \frac{1}{2}\int_0^{\pi} \frac{\sin t}{\sqrt{t + n\pi}}\, dt,$$

whence it is clear that u_n is positive and decreasing with increasing positive integral n. Moreover, it follows from the inequality

$$u_n < \frac{1}{2}\int_0^{\pi} \frac{dt}{\sqrt{n\pi}} = \frac{1}{2}\sqrt{\frac{\pi}{n}}$$

that $u_n \to 0$ as $n \to +\infty$. All this implies that the alternating series

$$u_0 - u_1 + u_2 - u_3 + \ldots + (-1)^n u_n + \ldots \tag{38}$$

is convergent [I, 123].

We now suppose that

$$\sqrt{m\pi} \leqslant b \leqslant \sqrt{(m+1)\pi}, \tag{39}$$

and consider the integral

$$\int_0^b \sin(x^2)\,dx = \int_0^{\sqrt{\pi}} \sin(x^2)\,dx + \int_{\sqrt{\pi}}^{\sqrt{2\pi}} \sin(x^2)\,dx + \ldots + \int_{\sqrt{(m-1)\pi}}^{\sqrt{m\pi}} \sin(x^2)\,dx +$$

$$+ \int_{\sqrt{m\pi}}^{b} \sin(x^2)\,dx = u_0 - u_1 + \ldots + (-1)^{m-1} u_{m-1} + \theta(-1)^m u_m, \tag{40}$$

where $0 \leqslant \theta < 1$, since the last interval $(\sqrt{m\pi}, b)$ consists of only part of $(\sqrt{m\pi}, \sqrt{(m+1)\pi})$ or is even absent with $b = \sqrt{m\pi}$. If $b \to \infty$ and the integral m, given by inequality (39), tends to $(+\infty)$, the existence of the improper integral

$$\int_0^{+\infty} \sin(x^2)\,dx = \lim_{b \to +\infty} \int_0^b \sin(x^2)\,dx = u_0 - u_1 + u_2 - u_3 + \ldots$$

follows from the convergence of series (38) and equation (40).

The existence of the improper integral in this case is due to the alternation of the integrand, and to the fact that the successive areas, above and below the x axis, decrease in size and tend to zero on moving away from the origin, this last being due to the indefinite compression of the areas and not to the fact that their heights tend to zero.

Integral (36) may be considered in exactly the same way.

We obtain the following value for integral (37) in Volume III:

$$\int_0^{+\infty} \sin(x^2)\,dx = \int_0^{+\infty} \cos(-x^2)\,dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}\,.$$

The integrals written are known as Fresnel or diffraction integrals, the latter name being due to the role they play in optics.

84. Uniformly convergent integrals.† If the integrand of an improper integral depends on a parameter y, the numbers η and N mentioned in the general tests 1 and 2 of [82] in general also depend on y.

† It may be useful to revise the theory of uniformly convergent series in Volume I before reading the present section.

If, as y varies in the interval $\alpha \leqslant y \leqslant \beta$, the numbers η and N in the conditions

$$\left|\int_{b-\varepsilon'}^{b-\varepsilon''} f(x,y)\,\mathrm{d}x\right| < \delta \quad \text{for} \quad 0 < \varepsilon' \quad \text{and} \quad \varepsilon'' < \eta \tag{41}$$

$$\left|\int_{b'}^{b''} f(x,y)\,\mathrm{d}x\right| < \delta \quad \text{for } b' \quad \text{and} \quad b'' > N \tag{42}$$

can be chosen independently of y, the improper integrals

$$\int_a^b f(x,y)\,\mathrm{d}x, \quad \int_a^{+\infty} f(x,y)\,\mathrm{d}x \tag{43}$$

are said to be uniformly convergent with respect to y.

In particular, the integrals encountered when applying Cauchy's tests are uniformly convergent if the constants A and p are independent of y.

Every convergent improper integral can be written in the form of a convergent series in which each term is an ordinary integral. We can use this approach for the above; let us take the first of integrals (43). Let

$$\varepsilon_1, \ \varepsilon_2, \ \varepsilon_3, \ldots, \ \varepsilon_n, \ldots \tag{44}$$

be a series of positive, decreasing terms which tend to zero; then we can write:

$$\int_a^b f(x,y)\,\mathrm{d}x = \int_a^{b-\varepsilon_1} f(x,y)\,\mathrm{d}x + \int_{b-\varepsilon_1}^{b-\varepsilon_2} + \int_{b-\varepsilon_2}^{b-\varepsilon_3} + \ldots +$$

$$+ \int_{b-\varepsilon_n}^{b-\varepsilon_{n+1}} + \ldots = u_0(y) + u_1(y) + u_2(y) + \ldots + u_n(y) + \ldots, \tag{45}$$

where

$$u_n(y) = \int_{b-\varepsilon_n}^{b-\varepsilon_{n+1}} f(x,y)\,\mathrm{d}x. \tag{46}$$

For the second of integrals (43), we can specify a series of indefinitely increasing numbers

$$b_1, \ b_2, \ b_3, \ldots, \ b_n, \ldots \tag{47}$$

and obtain:

$$\int_{a_n}^{+\infty} f(x,y)\,\mathrm{d}x = \int_a^{b_1} f(x,y)\,\mathrm{d}x + \int_{b_1}^{b_2} + \int_{b_2}^{b_3} + \ldots + \int_{b_n}^{b_{n+1}} + \ldots =$$

$$= u_0(y) + u_1(y) + u_2(y) + \ldots + u_n(y) + \ldots \tag{48}$$

It follows at once from the definition of uniform convergence of integrals and series [I, 143] that, if an improper integral is uniformly convergent, the series corresponding to it is also uniformly convergent for any choice of numbers (44) or (47). In fact, if we sum say remote terms of series (45), we get the integral over an interval close to b for which inequality (41) is satisfied.

Uniformly convergent integrals have similar properties to those of uniformly convergent series [I, 146]. We shall state these explicitly for the second of integrals (43), but what is said is likewise applicable to the first integral.

(1) *If $f(x, y)$ is continuous for $a \leqslant x$ and for y in a finite interval $\alpha \leqslant y \leqslant \beta$, and if*

$$\int_a^{+\infty} f(x,y)\,\mathrm{d}x \tag{49}$$

is uniformly convergent, the integral is a continuous function of y in $\alpha \leqslant y \leqslant \beta$.

(2) *With the same conditions, the formula for integration under the integral sign is also valid:*

$$\int_\alpha^\beta \mathrm{d}y \int_a^\infty f(x,y)\,\mathrm{d}x = \int_a^{+\infty} \mathrm{d}x \int_\alpha^\beta f(x,y)\mathrm{d}y. \tag{50}$$

(3) *If integral* (49) *is convergent with continuity of $f(x, y)$ and $\partial f(x, y)/\partial y$, whilst*

$$\int_a^{+\infty} \frac{\partial f(x,y)}{\partial y}\,\mathrm{d}x \tag{51}$$

is uniformly convergent, the formula for differentiation under the integral sign is valid:

$$\frac{\mathrm{d}}{\mathrm{d}y}\int_a^{+\infty} f(x,y)\,\mathrm{d}x = \int_a^{+\infty} \frac{\partial f(x,y)}{\partial y}\,\mathrm{d}x. \tag{52}$$

We give the proofs of properties (1) and (3) as examples. The terms of series (48):

$$u_n(y) = \int_{b_n}^{b_{n+1}} f(x,y)\,\mathrm{d}x, \tag{53}$$

are continuous by what was proved in [80], and the series is uniformly convergent due to the uniform convergence of the integral; hence the sum of the series, i.e. integral (49), is also continuous [I, 146].

We prove (3) by noticing the consequence of [80], that integral (53) can be differentiated under the integral sign, i.e.

$$u'_n(y) = \int_{b_n}^{b_{n+1}} \frac{\partial f(x, y)}{\partial y} \, dx .$$

But we have, by the uniform convergence of integral (51), the uniformly convergent series:

$$\int_a^{+\infty} \frac{\partial f(x, y)}{dy} \, dx = \sum_{n=0}^{\infty} \int_{b_n}^{b_{n+1}} \frac{\partial f(x, y)}{\partial y} \, dx = \sum_{n=0}^{\infty} u'_n(y) . \tag{54}$$

Series (48) is thus convergent, whilst the series of derivatives is uniformly convergent. Hence it follows [I, 146] that the sum of series (54) is the derivative of the sum of series (48), which leads us to (52).

We indicate a simple test for absolute and uniform convergence of improper integrals, similar to the corresponding test for series [I, 147]. To be explicit, we take the second of integrals (43); the test is similar for the first integral.

Let $f(x, y)$ be continuous as usual for $a \leqslant x$ and $\alpha \leqslant y \leqslant \beta$. If the positive function $\varphi(x)$ exists, continuous for $a \leqslant x$, such that $| f(x, y) | \leqslant \varphi(x)$ for $a \leqslant x$ and $\alpha \leqslant y \leqslant \beta$, and if

$$\int_a^{+\infty} \varphi(x) \, dx \tag{55}$$

is convergent, integral (49) *is absolutely and uniformly convergent (with respect to y).* By the convergence of (55), for any given $\delta > 0$ there exists an N such that:

$$\int_{b'}^{b''} \varphi(x) \, dx < \delta \qquad \text{for } b', \ b'' > N,$$

this N being independent of y, since $\varphi(x)$ does not contain y. But we have, since $| f(x, y) | \leqslant \varphi(x)$:

$$\left| \int_{b'}^{b''} f(x, y) \, dx \right| \leqslant \int_{b'}^{b''} | f(x, y) | \, dx \leqslant \int_{b'}^{b''} \varphi(x) \, dx < \delta \qquad \text{for } \ b', b'' > N,$$

i.e. the same N, independent of y, is suitable for integral (49) and further, for the integral

$$\int_a^{+\infty} | f(x, y) | \, dx,$$

which proves our statement.

85. Examples. 1. We consider Example 3 of [81] in more detail:

$$I\ (\alpha, \beta) = \int_0^\infty e^{-\alpha x} \frac{\sin \beta x}{x}\, dx. \tag{56}$$

Let α be a fixed positive number for the present, so that (56) depends on the parameter β. We notice that $(\sin \beta x)/x$ remains continuous at $x = 0$, where it takes the value β; hence (56) is an improper integral only on account of the infinite limit. For positive $x > 1$, we have $|(\sin \beta x)/x| < 1$ and therefore

$$\left| e^{-\alpha x} \frac{\sin \beta x}{x} \right| < e^{-\alpha x},$$

whilst the integral

$$\int_1^\infty e^{-\alpha x}\, dx = \left[-\frac{1}{\alpha}\, e^{-\alpha x}\right]_{x=1}^{x=\infty} = \frac{1}{\alpha}\, e^{-\alpha}$$

is convergent; hence, by the test proved, (56) is uniformly convergent with respect to β. On differentiating with respect to β under the integral sign, we get

$$\int_0^\infty e^{-\alpha x} \cos \beta x\, dx,$$

which is also uniformly convergent, since $|e^{-\alpha x} \cos \beta x| < e^{-\alpha x}$. Hence it follows that (56) is a continuous function of β and can be differentiated under the integral sign. All the arguments of the example mentioned will be justified if it is proved that $\lim\limits_{\alpha \to +0} I(\alpha, \beta) = I(0, \beta)$, i.e. (56) with fixed β is a continuous function of α to the right of zero. We prove the continuity for $\alpha \geqslant 0$. The convergence for $\alpha = 0$ has already been shown above.

We can take $\beta > 0$ without loss of generality, since $\beta < 0$ reduces to this case by simply changing the sign of the integral, whilst our statement is obvious if $\beta = 0$.

We shall proceed as in [83] for Fresnel integrals. We subdivide $(0, +\infty)$ as follows:

$$\left(0, \frac{\pi}{\beta}\right), \left(\frac{\pi}{\beta}, \frac{2\pi}{\beta}\right), \ldots, \left(\frac{n\pi}{\beta}, \frac{(n+1)\pi}{\beta}\right), \ldots$$

so that the integrand

$$f(x) = e^{-\alpha x} \frac{\sin \beta x}{x}\ (\alpha \geqslant 0 \text{ and } \beta > 0)$$

is $(+)$ in the first sub-interval, $(-)$ in the second, and so on. We write

$$u_n(\alpha) = (-1)^n \int_{\frac{n\pi}{\beta}}^{\frac{(n+1)\pi}{\beta}} e^{-\alpha x} \frac{\sin \beta x}{x}\, dx.$$

We replace x by a new variable $t : t = x - n\pi/\beta$, and get:

$$u_n(\alpha) = \int_0^{\frac{\pi}{\beta}} e^{-\alpha t - \frac{n\alpha\pi}{\beta}} \frac{\sin \beta t}{t + \frac{n\pi}{\beta}} \, dt,$$

whence it is evident that the $u_n(\alpha)$ are positive and decrease with increasing n.

Furthermore, since

$$|u_n(\alpha)| < \int_0^{\frac{\pi}{\beta}} \frac{1}{\frac{n\pi}{\beta}} \, dt = \frac{1}{n} \tag{57}$$

it follows that $u_n(\alpha) \to 0$ as $n \to +\infty$.

We can thus write our integral for $\alpha \geqslant 0$ as the sum of an alternating series:

$$\int_0^{+\infty} e^{-\alpha x} \frac{\sin \beta x}{x} \, dx = u_0(\alpha) - u_1(\alpha) + u_2(\alpha) - \ldots + (-1)^n u_n(\alpha) + \ldots \tag{58}$$

We have for the remainder term of the series, by (57) and the theorem of [I, 123]:

$$|r_n(\alpha)| < |u_{n+1}(\alpha)| < \frac{1}{n+1},$$

where $1/(n+1) \to 0$ as $n \to +\infty$ independently of α. The series is thus uniformly convergent for $\alpha \geqslant 0$, and its sum is therefore continuous [I, 146], inasmuch as the terms $u_n(\alpha)$ are continuous by [80].

We remark that the uniform convergence of the integral does not follow without further argument from the uniform convergence alone of series (58) for $\alpha \geqslant 0$. It can be shown that the integral in this particular case is in fact uniformly convergent for $\alpha \geqslant 0$.

We notice that

$$\int_0^{\infty} \frac{\sin \beta x}{x} \, dx,$$

which is equal to $\pi/2$ for $\beta > 0$, $(-\pi/2)$ for $\beta < 0$, and zero for $\beta = 0$, gives a function of β with a break in continuity at $\beta = 0$. Hence it follows that this integral cannot converge uniformly with respect to β in an interval containing $\beta = 0$. If we take the interval to the right of zero, the integral, equal to $\pi/2$, has a zero derivative with respect to β; the integral cannot be differentiated with respect to β under the integral sign, however, since this yields the integral of $\cos \beta x$ from 0 to ∞, which is meaningless.

2. In Example 4 of [81], we differentiated

$$\int_0^{\infty} e^{-\alpha x} \, dx = \frac{1}{\alpha} \qquad (\alpha > 0)$$

k times with respect to α under the integral sign. To justify this operation, it is sufficient to show that, with positive integral k,

$$\int_0^\infty e^{-\alpha x} x^k \,dx$$

is uniformly convergent for every interval $c \leqslant \alpha \leqslant d$, where $c > 0$. Since $x \geqslant 0$ in the interval of integration, obviously $e^{-\alpha x} \leqslant e^{-cx}$ and $e^{-\alpha x} x^k \leqslant \leqslant e^{-cx} x^k$, and by the test for uniform convergence proved in [84], it is sufficient for us to prove the convergence of

$$\int_0^\infty e^{-cx} x^k \,dx.$$

We write $f(x) = e^{-cx} x^k$ and apply l'Hôpital's rule in the ordinary form [I, 65] to see that $f(x)\, x^2 = e^{-cx} x^{k+2} \to 0$ as $x \to +\infty$; hence, by the test proved in [82], the integral written is in fact convergent.

3. We obtained the solution of Abel's problem in [79] as:

$$u(z) = \frac{\sqrt{2g}}{\pi} \frac{d}{dz} \int_0^z \frac{\varphi(h)\,dh}{\sqrt{z-h}}\,.$$

We show how the derivative on the right can be evaluated. We write:

$$I(z) = \int_0^z \frac{\varphi(h)\,dh}{\sqrt{z-h}}\,.$$

If we were to differentiate with respect to z under the integral sign, we should get $(z-h)^{-3/2}$ in the integrand, which leads to a divergent integral [82], so that we must proceed otherwise. We assume the existence of a continuous, bounded derivative $\varphi'(h)$ in the neighbourhood of $h = 0$ with $h > 0$, and integrate $I(z)$ by parts:

$$\int_0^z \frac{\varphi(h)\,dh}{\sqrt{z-h}} = -2\int_0^z \varphi(h)\,d\sqrt{z-h} = -2\varphi(h)\sqrt{z-h}\,\Big|_{h=+0}^{h=z} +$$

$$+ 2\int_0^z \varphi'(h)\sqrt{z-h}\,dh = 2\varphi(+0)\sqrt{z} + 2\int_0^z \varphi'(h)\sqrt{z-h}\,dh.$$

It may be recalled that $\varphi(+0) = \lim\limits_{h\to+0} \varphi(h)$. This is a constant which in general, differs from zero, whereas $\varphi(0) = 0$ by definition. We obtain on differentiating the above expression and using (21) of [80]:

$$\frac{d}{dz}\int_0^z \frac{\varphi(h)\,dh}{\sqrt{z-h}} = \frac{\varphi(+0)}{\sqrt{z}} + \int_0^z \frac{\varphi'(h)}{\sqrt{z-h}}\,dh. \qquad (59)$$

If $\varphi(h)$ is constant, $\varphi'(h) = 0$, and we arrive at the expression already obtained, If $\varphi(+0) = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}z}\int_0^z \frac{\varphi(h)\,\mathrm{d}h}{\sqrt{z-h}} = \int_0^z \frac{\varphi'(h)\,\mathrm{d}u}{\sqrt{z-h}} \tag{59_1}$$

We pass over the proof of the applicability of (21) of [80] to the improper integral $I(z)$. We notice that, if h is replaced by a new variable of integration u, where $h = zu$, an integral with constant limits is obtained for $I(z)$:

$$I(z) = \sqrt{z}\int_0^1 \frac{\varphi(zu)\,\mathrm{d}u}{\sqrt{1-u}}\,.$$

It is easily shown that, with the above assumption of the existence of a continuous, bounded derivative $\varphi'(h)$ for $h > 0$, we can differentiate under the integral sign:

$$\frac{\mathrm{d}I(z)}{\mathrm{d}z} = \frac{1}{2\sqrt{z}}\int_0^1 \frac{\varphi'(zu)\,\mathrm{d}u}{\sqrt{1-u}} + \sqrt{z}\int_0^1 \frac{\varphi'(zu)\,u\,\mathrm{d}u}{\sqrt{1-u}}\,.$$

If we integrate the first term by parts and return to the former variable h, we again get (59).

86. Improper multiple integrals. We shall start by considering improper double integrals. Improper integrals can be of two kinds, as above: either the integrand, or the domain of integration, is unbounded. We begin with the first kind, and let $f(M)$ be continuous in a finite domain (σ) excluding the point C, in the vicinity of which $f(M)$ is unbounded. We surround C with a small domain (Δ), so that $f(M)$ is continuous throughout the remaining domain $(\sigma - \Delta)$, and

$$\iint_{(\sigma-\Delta)} f(M)\,\mathrm{d}\sigma.$$

has a meaning.

If the integral tends to a definite limit as Δ contracts to C, independently of how the contraction takes place, the limit is called the improper integral of $f(M)$ over (σ):

$$\iint_{(\sigma)} f(M)\,\mathrm{d}\sigma = \lim \iint_{(\sigma-\Delta)} f(M)\,\mathrm{d}\sigma. \tag{60}$$

To start with, let $f(M)$ be positive, or more precisely, non-negative in the neighbourhood of C. Let (Δ') and (Δ'') be two small domains, where (Δ'') lies inside (Δ'). The integral over $(\sigma - \Delta'')$ will now differ from the integral over $(\sigma - \Delta')$ by a positive quantity, equal to the integral over $(\Delta' - \Delta'')$ in which $f(M) \geqslant 0$. It follows at once from

this that integral (60) increases on indefinite contraction of the (Δ) to C (if each successive (Δ) is included in the previous (Δ)); hence, the integral either tends to a limit or increases indefinitely. If the limit is finite for one method of contracting (Δ) to C, the same limit will be found by any other method of contraction. The existence of a limit is characterized by the fact that the integral over any domain which excludes C but lies in the neighbourhood of C, where $f(M)$ is positive, remains less than a definite positive number (the integral here will tend to zero if the neighbourhood contracts to C). If $f(M) \leqslant 0$ near C, we get the above case on taking the minus sign outside the integral. We now let $f(M)$ change sign in any small neighbourhood of C. We shall consider in this case only absolutely convergent integrals, i.e. such that

$$\iint\limits_{(\sigma)} |f(M)|\,\mathrm{d}\sigma \tag{61}$$

has a meaning, that is to say, is convergent. The integrand is now non-negative, and the above remarks are applicable. It follows in particular from these remarks that, if $f_1(M)$, $f_2(M)$ are two positive functions, whilst $f_1(M) \leqslant f_2(M)$ and the integral of $f_2(M)$ is convergent, the integral of $f_1(M)$ is likewise convergent. We now write our $f(M)$ as the difference between two positive functions: $f(M) = | f(M) | - [| f(M) | - f(M)]$. Integral (61) is convergent by hypothesis, so that the integral of $2 | f(M) |$ is convergent. We have $[| f(M) | - f(M)] = 2 | f(M) |$ when $f(M) \leqslant 0$, and $= 0$ when $f(M) > 0$, i.e. $0 \leqslant | f(M) | - f(M) \leqslant 2 | f(M) |$, so that the integral of this positive function is also convergent. It follows that the integral of the difference: $| f(M) | - [| f(M) | - f(M)] = f(M)$ is convergent. Hence, *if integral* (61) *is convergent, the integral of* $f(M)$ *is also convergent.*

We indicate a sufficient condition for the convergence of (61): *if* $| f(M) | \leqslant A/r^p$ *in the neighbourhood of* C, *where* r *is the distance of the variable point* M *from* C, *and* A *and* p *are constants and* $p < 2$, *integral* (61) *is convergent.* From what has been said above, it is enough for us to show that (61) is bounded over any domain (σ') which excludes C but lies in a circle of centre C, radius r_0. We use polar coordinates with C as origin, and note the inequality written above for $| f(M) |$; we have:

$$\iint\limits_{(\sigma')} |f(M)|\,\mathrm{d}\sigma \leqslant A \iint\limits_{(\sigma')} \frac{1}{r^p}\, r\,\mathrm{d}r\,\mathrm{d}\varphi = A \iint\limits_{(\sigma')} \frac{1}{r^{p-1}}\,\mathrm{d}r\,\mathrm{d}\varphi.$$

The domain (σ') must lie inside a ring bounded by the circles $r = \eta$ and $r = r_0$, where η can be taken as small as desired. The integrand is positive, so that we can only increase the result by integrating over the whole of the ring, i.e.

$$\iint\limits_{(\sigma)} |f(M)|\,\mathrm{d}\sigma \leqslant A\int\limits_0^{2\pi}\mathrm{d}\varphi\int\limits_\eta^{r_0}\frac{1}{r^{p-1}}\,\mathrm{d}r = \frac{2\pi A}{2-p}(r_0^{2-p} - \eta^{2-p}).$$

Since $2 - p > 0$, this gives us finally for the integral over (σ'):

$$\iint\limits_{(\sigma')} |f(M)|\,\mathrm{d}\sigma \leqslant \frac{2\pi A}{2-p}r_0^{2-p}, \tag{62}$$

which proves the above statement.

It may be noted that the integral over (σ') may be as small as desired, given a sufficiently small r_0.

The definition is similar for an improper triple integral over a finite domain (v), when $f(M)$ is unbounded near the point C. All the above remarks apply, except that the sufficient condition for absolute convergence now reads: *if* $|f(M)| \leqslant A/r^p$ *in the neighbourhood of the point* C, *where* r *is the distance of the variable point* M *from* C, *and* A *and* p *are constants and* $p < 3$, *the integral*

$$\iiint\limits_{(v)} f(M)\,\mathrm{d}v$$

is absolutely convergent. Here, the condition $p < 2$ is replaced by $p < 3$, since an elementary volume in polar coordinates in space is $\mathrm{d}v = r^2 \sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\varphi$ (with r^2 instead of the r in $\mathrm{d}\sigma = r\,\mathrm{d}r\,\mathrm{d}\varphi$).

We now take the case when the domain of integration (σ) is unbounded, i.e. extends to infinity in every direction. Let (σ_1) be a finite domain contained in (σ), and let it be extended so that any point M of (σ) eventually lies inside (σ_1). Assuming that $f(M)$ is continuous in (σ), we can form the integral

$$\iint\limits_{(\sigma_1)} f(M)\,\mathrm{d}\sigma. \tag{63}$$

If the integral tends to a definite limit on indefinite extension of (σ_1), the limit is defined as the integral of $f(M)$ over the infinite domain (σ):

$$\iint\limits_{(\sigma)} f(M)\,\mathrm{d}\sigma = \lim \iint\limits_{(\sigma_1)} f(M)\,\mathrm{d}\sigma. \tag{64}$$

If $f(M) \geqslant 0$ for all sufficiently remote M, integral (63) either has a limit or increases indefinitely on extension of (σ_1). The first case is

characterized by the fact that the integral is bounded over any domain, or in fact any finite number of domains, belonging to (σ) and lying outside a circle with centre at the origin and radius r_0 (with this, the integral tends to zero if $r_0 \to \infty$). Let (σ') denote the set of these domains. We notice a consequence of the definition of improper integral, that if

$$\iint\limits_{(\sigma)} |f(M)|\, d\sigma, \tag{65}$$

is convergent, (64) is also convergent. Integral (64) is now said to be absolutely convergent, and we shall only consider integrals of this type. The following sufficient condition for convergence may easily be proved: *if* $|f(M)| \leqslant A/r^p$ *for all sufficiently remote points* M, *where* r *is the distance from any fixed point (origin) to the variable point* M, *and* A *and* p *are constants and* $p > 2$, *integral* (64) *is convergent.* If we use the inequality written and introduce polar coordinates, we have:

$$\iint\limits_{(\sigma')} |f(M)|\, d\sigma \leqslant A \iint\limits_{(\sigma')} \frac{1}{r^{p-1}}\, dr\, d\varphi.$$

The set (σ') must be contained within a ring bounded by circles of radii $r = r_0$, $r = R$, where R can be as large as desired. Integration over the whole of the ring gives:

$$\iint\limits_{(\sigma')} |f(M)|\, d\sigma \leqslant A \int\limits_0^{2\pi} d\varphi \int\limits_{r_0}^{R} \frac{1}{r^{p-1}}\, dr = \frac{2\pi A}{p-2}\left(\frac{1}{r_0^{p-2}} - \frac{1}{R^{p-2}}\right).$$

Since $p - 2 > 0$, we have finally for the integral over (σ'):

$$\iint\limits_{(\sigma')} |f(M)|\, d\sigma \leqslant \frac{2\pi A}{p-2}\frac{1}{r_0^{p-2}},$$

which proves our proposition. The integral over (σ') may be as small as desired, with r_0 sufficiently large.

An improper triple integral over an infinite domain is similarly defined. The condition $p > 2$ in the last theorem has to be replaced by $p > 3$ for triple integrals. It may also be noted that the above remarks about improper double integrals in which $f(M)$ tends to infinity are applicable to improper integrals over curved surfaces; these reduce to plane surface integrals, as we have seen [63].

An absolutely convergent improper integral reduces, as we saw above, to the integrals of the non-negative functions $|f(M)|$ and $[|f(M)| - f(M)]$, whilst it is of no consequence for these integrals

how (Δ) contracts to the point C or how (σ_1) is extended. It can always be assumed that (Δ) is a circle or a sphere (Δ_ϱ) with centre C and whose radius ϱ tends to zero, and that (σ_1) is a part of (σ) consisting of a circle (K_R) with centre at the origin, the radius of which increases indefinitely. The uniform convergence of improper multiple integrals which depend on a parameter may easily be defined with the aid of these remarks. For instance, *integral* (60), *when its integrand depends on the parameter* a, *is said to be uniformly convergent with respect to* a *if, for any positive* δ, *there exists a positive* η, *independent of* a, *such that*

$$\left|\iint_{(\sigma')} f(M)\,\mathrm{d}\sigma\right| < \delta,$$

where (σ') *is any part of* (σ) *contained in the circle* (Δ_η). The uniform convergence of the other improper integrals is similarly defined. In particular, it follows from (62) that the integral is absolutely and uniformly convergent if A and p are independent of a.

The properties and test of [84] apply for the uniform convergence of multiple integrals.

A more difficult problem arises in the case of improper multiple integrals where the integrand is unbounded in the neighbourhood of a line (l) instead of in the neighbourhood of a point. Here, the line has to be excluded with the aid of a domain (Δ), then (Δ) is allowed to contract to the line.

If $f(M)$ is assumed positive in the neighbourhood of (l), we can say that the integral over the remaining domain either tends to a limit or to infinity, independently of the method of contraction of (Δ) to (l). Definitions similar to the above are obtained for absolutely convergent integrals, which are the only type that we take into consideration.

87. Examples. 1. We consider:

$$\iint_{(\sigma)} \frac{\mathrm{d}x\,\mathrm{d}y}{(1+x^2+y^2)^a} \qquad (a \neq 1),$$

where (σ) is the whole plane. We use polar coordinates and integrate over the circle (K_R) with centre at the origin and radius R, and get

$$\iint_{(K_R)} \frac{r\,\mathrm{d}r\,\mathrm{d}\varphi}{(1+r^2)^a} = \frac{\pi}{1-a}\left[\frac{1}{(1+R^2)^{a-1}} - 1\right].$$

If $a < 1$, the right-hand side increases indefinitely on indefinite increase of R and the integral is divergent. If $a > 1$, the right-hand side has a finite limit

$\pi/(a-1)$, so that the integral is convergent and equal to $\pi/(a-1)$. The convergence in this second case can be proved by applying the sufficient condition stated in the previous article.

2. We take:

$$\iint\limits_{(\sigma)} \frac{y\,\mathrm{d}x\,\mathrm{d}y}{\sqrt{x}},$$

where (σ) is the square bounded by $x=0$, $x=1$, $y=0$, $y=1$. The integrand tends to infinity along the side $x=0$. We exclude this side by means of a narrow vertical strip, and integrate over the rectangle (σ_ε), bounded by $x=\varepsilon$, $x=1$, $y=0$, $y=1$ $(\varepsilon>0)$:

$$\iint\limits_{(\sigma_\varepsilon)} \frac{y\,\mathrm{d}x\,\mathrm{d}y}{\sqrt{x}} = \int_0^1 y\,\mathrm{d}y \int_\varepsilon^1 \frac{\mathrm{d}x}{\sqrt{x}} = 1-\sqrt{\varepsilon},$$

and we have the limit unity as $\varepsilon \to 0$, i.e. the integral is convergent and equal to unity.

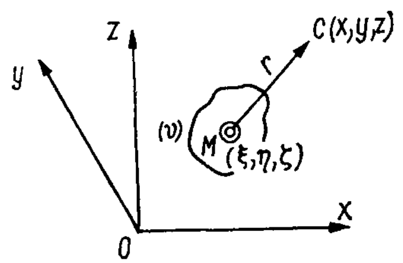

FIG. 78

3. *The attraction exerted by a mass on a particle, located outside or inside the mass* (Fig. 78). Let the particle, at $C(x, y, z)$, be of unit mass. We divide the attracting body (v) into elementary masses Δm containing the points $M(\xi, \eta, \zeta)$. We take the total mass Δm as concentrated at M and get the approximation for the attraction at C due to Δm:

$$\frac{\Delta m}{r^2},$$

where r is the distance CM, and the gravitational constant is taken equal unity. Since the attraction is along CM, its projections on the axes are:

$$\frac{\Delta m}{r^2}\cdot\frac{\xi-x}{r};\quad \frac{\Delta m}{r^2}\cdot\frac{\eta-y}{r};\quad \frac{\Delta m}{r^2}\cdot\frac{\zeta-z}{r}.$$

The projections of the total attraction will be given approximately by

$$X \sim \sum \frac{\xi-x}{r^3}\Delta m;\quad Y \sim \sum \frac{\eta-y}{r^3}\Delta m;\quad Z \sim \sum \frac{\zeta-z}{r^3}\Delta m.$$

If $\mu(\xi, \eta, \zeta)$ denotes the density of the body at M, we have:

$$\Delta m \sim \mu \Delta v,$$

and finally, if we decrease each element indefinitely whilst increasing their number, we find:

$$X = \iiint\limits_{(v)} \mu\frac{\xi-x}{r^3}\,\mathrm{d}v;\quad Y = \iiint\limits_{(v)} \mu\frac{\eta-y}{r^3}\,\mathrm{d}v;\quad Z = \iiint\limits_{(v)} \mu\frac{\zeta-z}{r^3}\,\mathrm{d}v. \qquad (66)$$

We notice that the coordinates (ξ, η, ζ) of the variable point M of (v) are the variables of integration here, the density $\mu(\xi, \eta, \zeta)$ being a function of them.

The coordinates (x, y, z) of C appear both directly in the numerator of the integrand and indirectly via

$$r = \sqrt{(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2},$$

and these coordinates act as parameters, X, Y and Z being functions of (x, y, z).

If C is outside the body, r never vanishes and we have ordinary integrals to deal with. If C is inside the body, the integrands in (66) tend to infinity as the point of integration M approaches C, so that we are concerned with improper integrals. These clearly have a meaning, however, if we take μ to be a continuous function, since, on writing μ_0 for the upper limit of the function $|\mu|$, we have:

$$\left|\mu \frac{\xi - x}{r^3}\right| = \left|\mu \frac{1}{r^2} \frac{\xi - x}{r}\right| < \frac{\mu_0}{r^2}; \quad \left|\mu \frac{\eta - y}{r^3}\right| < \frac{\mu_0}{r^2}; \quad \left|\mu \frac{\zeta - z}{r^2}\right| < \frac{\mu_0}{r^2}, \tag{67}$$

so that $p = 2$ in the rule given above and $A = \mu_0$.

A meaning may also be attached, all the more, to

$$U = \iiint\limits_{(v)} \frac{\eta \, \mathrm{d}v}{r}, \tag{68}$$

denoting the *potential* of the mass concerned at the point C. (A more detailed treatment of potential will be found below).

4. We have the obvious expressions

$$\frac{\xi - x}{r} = -\frac{\partial r}{\partial x}; \quad \frac{\eta - y}{r} = -\frac{\partial r}{\partial y}; \quad \frac{\zeta - z}{r} = -\frac{\partial r}{\partial z};$$

$$\frac{\xi - x}{r^3} = \left(-\frac{1}{r^2}\right)\left(-\frac{\xi - x}{r}\right) = \frac{\partial}{\partial x}\left(\frac{1}{r}\right);$$

$$\frac{\eta - y}{r^3} = \frac{\partial}{\partial y}\left(\frac{1}{r}\right); \quad \frac{\zeta - z}{r^3} = \frac{\partial}{\partial r}\left(\frac{1}{r}\right),$$

so that integrals (66) can be written in the form:

$$X = \iiint\limits_{(v)} \mu \frac{\partial}{\partial x}\left(\frac{1}{r}\right) \mathrm{d}v; \qquad Y = \iiint\limits_{(v)} \mu \frac{\partial}{\partial y}\left(\frac{1}{r}\right) \mathrm{d}v;$$

$$Z = \iiint\limits_{(v)} \mu \frac{\partial}{\partial z}\left(\frac{1}{r}\right) \mathrm{d}v,$$

i.e. these integrals are found by differentiating (68) under the integral sign with respect to x, y and z. The differentiation is carried out with respect to the coordinates (x, y, z) of points at which the integrand is discontinuous, and this case does not come within the scope of the conditions for the theorems established above [84], regarding continuity and the possibility of differentiation under the integral sign. We shall see later [200] that, provided $\mu(\xi, \eta, \zeta)$ is continuous, integrals X, Y, Z are continuous functions of (x, y, z) throughout space, whilst U is a continuous function with continuous first order partial

derivatives, these latter being obtainable by differentiation of (68) under the integral sign, i.e.

$$X = \frac{\partial u}{\partial x}; \quad Y = \frac{\partial u}{\partial y}; \quad Z = \frac{\partial u}{\partial z}.$$

Further differentiation of the potential U with respect to x, y, z under the integral sign gives us, since $\mu(\xi, \eta, \zeta)$ is independent of (x, y, z):

$$\left.\begin{array}{c} \dfrac{\partial^2 U}{\partial x^2} = \iiint\limits_{(v)} \mu \dfrac{\partial^2}{\partial x^2}\left(\dfrac{1}{r}\right) \mathrm{d}v; \quad \dfrac{\partial^2 U}{\partial y^2} = \iiint\limits_{(v)} \mu \dfrac{\partial^2}{\partial y^2}\left(\dfrac{1}{r}\right) \mathrm{d}v; \\ \dfrac{\partial^2 U}{\partial z^2} = \iiint\limits_{(v)} \mu \dfrac{\partial^2}{\partial z^2}\left(\dfrac{1}{r}\right) \mathrm{d}v. \end{array}\right\} \quad (69)$$

These formulae are valid only when $C(x, y, z)$ lies outside the attracting body (v), in which case they contain only proper integrals. If C is inside (v), it is easily shown by direct double differentiation that

$$\left.\begin{array}{c} \dfrac{\partial^2}{\partial x^2}\left(\dfrac{1}{r}\right) = \dfrac{3(\xi - x)^2}{r^5} - \dfrac{1}{r^3}; \quad \dfrac{\partial^2}{\partial y^2}\left(\dfrac{1}{r}\right) = \dfrac{3(\eta - y)^2}{r^5} - \dfrac{1}{r^3}; \\ \dfrac{\partial^2}{\partial z^2}\left(\dfrac{1}{r}\right) = \dfrac{3(\zeta - z)^2}{r^5} - \dfrac{1}{r^3}, \end{array}\right\} \quad (70)$$

and the convergence test of [87] is no longer applicable to (69), i.e. with C inside (v), the second derivatives of U cannot be obtained by differentiating twice under the integral sign.

We get, on adding equations (70):

$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{r}\right) + \frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right) + \frac{\partial^2}{\partial z^2}\left(\frac{1}{r}\right) = \frac{3[(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2]}{r^5} - \frac{3}{r^3} = 0,$$

and therefore, on adding equations (69), which are valid for C outside (v), we find the equation:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0. \quad (71)$$

The mass potential $U(x, y, z)$ *of a volume satisfies* (71) *at points* $C(x, y, z)$ *lying outside the volume*. We show further on how the equation has to be modified if C lies inside the volume.

5. We take the case of a homogeneous sphere of radius a (μ = constant). We take the axis OZ along OC, where O is the centre of the sphere (Fig. 79), and introduce spherical coordinates $(\varrho, \theta, \varphi)$:

$$U = \iiint\limits_{(v)} \mu \frac{\mathrm{d}v}{r} = \mu \int_0^{2\pi}\int_0^{\pi}\int_0^{a} \frac{\varrho^2 \sin\theta}{r} \,\mathrm{d}\varphi \,\mathrm{d}\theta \,\mathrm{d}\varrho. \quad (72)$$

Obviously,

$$r^2 = \varrho^2 + z^2 - 2\varrho z \cos\theta. \tag{73}$$

We take first the integration with respect to θ:

$$\int_0^\pi \frac{\sin\theta \, d\theta}{r}.$$

We take r as variable instead of θ, ϱ and φ being assumed constant. Two cases must be distinguished here: if $z > \varrho$, as θ varies from 0 to π with ϱ and

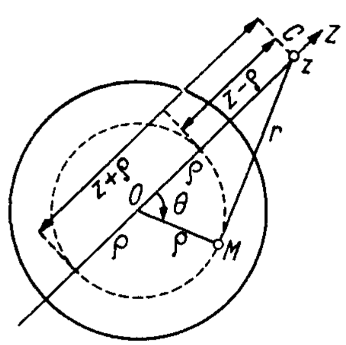

FIG. 79

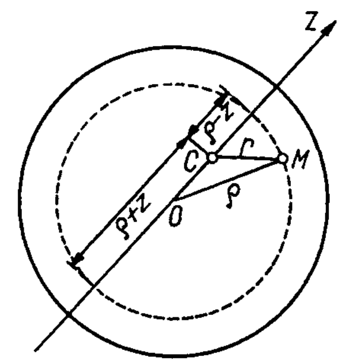

FIG. 80

φ constant, r varies from $(z - \varrho)$ to $(z + \varrho)$; if $z < \varrho$, r varies from $(\varrho - z)$ to $(\varrho + z)$ (Fig. 80). Also, from (73), with constant ϱ and φ:

$$r \, dr = \varrho z \sin\theta \, d\theta; \quad \frac{\sin\theta \, d\theta}{r} = \frac{dr}{\varrho z}.$$

Hence we have:

$$\int_0^\pi \frac{\sin\theta \, d\theta}{r} = \begin{cases} \displaystyle\int_{z-\varrho}^{z+\varrho} \frac{dr}{\varrho z} = \frac{2}{z} & (z > \varrho) \\ \displaystyle\int_{\varrho-z}^{\varrho+z} \frac{dz}{\varrho z} = \frac{2}{\varrho} & (z < \varrho). \end{cases}$$

We again distinguish two cases, on substituting from here into (72):

(1) C is outside or on the surface of the sphere: we have $a \leqslant z$, and $\varrho \leqslant z$ throughout the interval $(0, a)$, so that

$$U = \mu \int_0^{2\pi} d\varphi \int_0^a \frac{2\varrho^2 \, d\varrho}{z} = \frac{4\pi a^3 \mu}{3z} = \frac{m}{z}, \tag{74}$$

where m is the total mass of the sphere.

(2) C lies inside the sphere (Fig. 80): here we have to divide $(0, a)$ into two intervals $(0, z)$ and (z, a), and we get

$$U = \mu \int_0^{2\pi} d\varphi \left[\int_0^z \frac{2\varrho^2 \, d\varrho}{z} + \int_z^a \frac{2\varrho^2 \, d\varrho}{\varrho} \right] = 2\pi\mu \left(a^2 - \frac{1}{3} z^2 \right); \qquad (75)$$

with $z = a$, i.e. the point on the surface of the sphere, (74) and (75) both give the same value for U, which proves the continuity of the function U.

We now calculate the attraction. This must be along OZ by symmetry, so that we only have to find:

$$Z = \frac{\partial U}{\partial z} .$$

We use (74) when C is outside the sphere:

$$Z = -\frac{m}{z^2}; \qquad (76)$$

when C is inside, (75) gives

$$Z = -\frac{4}{3} \pi\mu z. \qquad (77)$$

Expressions (76) and (77) coincide for $z = a$, proving the continuity of the attraction Z.

We see from (74), (76), (77) that the *potential and attraction of a homogeneous sphere at a point outside it are found by concentrating the total mass of the sphere at its centre. Further, the attraction at an internal point of the sphere is proportional to the distance of the point from the centre of the sphere.*

We chose the axes in a particular way for simplicity, with C lying on OZ, with the result that z in the above expressions is the distance of C from the centre of the sphere. With any other disposition of the axes (with the origin at the centre of the sphere), z must be replaced by $\sqrt{x^2 + y^2 + z^2}$, where $(x. y, z)$ are the coordinates of C as usual; (74) and (75) now become:

$$U = \frac{m}{\sqrt{x^2 + y^2 + z^2}} \text{ (} C \text{ outside the sphere);}$$

$$U = 2\pi\mu \left[a^2 - \frac{1}{3} (x^2 + y^2 + z^2) \right] \text{ (} C \text{ inside the sphere).}$$

The first expression for U evidently satisfies (71). If we differentiate the second expression twice with respect to x, y, z we get:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = -4\pi\mu \text{ (} C \text{ inside the sphere).} \qquad (78)$$

As we shall see later, this equation is valid for any volume (v) of variable density, when C lies inside (v).

6. Let the attracting mass be distributed over a surface (S) with surface density $\mu(M)$, which is a function of the position of the variable point M on (S). Let $C(x, y, z)$ be the attracted particle of unit mass as above, and r the distance $|CM|$; we get for the potential:

$$U = \iint\limits_{(S)} \frac{\mu(M)}{r} \, \mathrm{d}S \tag{79}$$

and for the projections of the attraction:

$$X = \frac{\partial U}{\partial x} = \iint\limits_{(S)} \mu(M) \frac{\partial}{\partial z}\left(\frac{1}{r}\right) \mathrm{d}S; \quad Y = \frac{\partial U}{\partial y} = \iint\limits_{(S)} \mu\,(M) \frac{\partial}{\partial y}\left(\frac{1}{r}\right) \mathrm{d}S;$$

$$Z = \frac{\partial U}{\partial z} = \iint\limits_{(S)} \mu(M) \frac{\partial}{\partial z}\left(\frac{1}{r}\right) \mathrm{d}S.$$

We generally speak of (79) as the *potential of a simple layer*. We only take into account in this problem the case when C lies outside (S), so that all the integrals are proper. Potential (79) now satisfies equation (71), as above.

§ 9. Supplementary remarks on the theory of multiple integrals

88. Preliminary concepts. We took as our starting point in the treatment of multiple integrals the intuitive ideas of area and volume. The present article is devoted to the justification of these ideas and to a rigorous discussion of the basis of the theory of such integrals. We start by establishing certain concepts and proving a number of useful theorems regarding sets of points. Our treatment is confined to the case of a plane, since all the arguments are easily extended to the case of space.

We take a plane referred to Cartesian axes XY. The *ε-neighbourhood of the point M* is defined as the circle with centre M and radius ε. We shall consider all possible *sets of points* in the plane, which can consist of either a finite or an infinite number of points. Let (P) be a given set of points. The point M is called a *limit-point of the set* (P) if an infinite set of points of (P) belongs to any ε-neighbourhood of M. The point M may belong to (P), or it may not. If all the limit-points of (P) belong to (P), we say that (P) is a *closed set*. M, belonging to (P), is called an interior point of (P) if all the points of an ε-neighbourhood of M belong to (P).

For example, let (P) be the set of all points lying inside the square: $0 < x < 1$, $0 < y < 1$. Every point is now an interior point of the set, and is also a limit-point. All the points belonging to the boundaries of the square, i.e. its sides, will also be limit-points of the set. Since we have not included these latter in (P), the set is not closed.

An *open set* or *domain* is defined as a set, all the points of which are interior points. A *connected domain* is defined as an open set (P), such that any two

points of (P) can be joined by a line, all the points of which belong to (P). With this definition, the interior points of a square form a connected domain, whereas the interior points of two separate squares do not. What we have referred to above as simply a domain is sometimes called an open domain. The *boundary of a domain* (P) is defined as the set (l) of points M' with this property: the M' do not themselves belong to (P), but any ε-neighbourhood of M' contains points of (P). Since (P) consists of interior points, we can say that an infinite set of points of (P) lies in any ε-neighbourhood of M', and we can define the contour (l) of a domain as the set of limit-points of (P) which do not belong to (P). It may easily be seen that (l) *is a closed set:* we show that, if N is a limit-point of (l), it belongs to (l). There are points M' of (l) in any ε-neighbourhood of N, by the definition of a limit-point, whilst N cannot belong to (P) since all the points of (P) are interior points. But there are points of (P) in any ε-neighbourhood of an M' (by the definition of boundary), and hence, there are points of (P) in any ε-neighbourhood of N, i.e. N in fact belongs to (l). If we include in (P) its boundary (l), a closed set $(\overline{P})$ is obtained, which is occasionally referred to as a *closed domain.* If there are points of $(\overline{P})$ in any ε-neighbourhood of a point M, whilst M does not belong to (P), there must be points of (l) in any ε-neighbourhood of M and it follows, since (l) is closed, that M belongs to (l) and therefore to $(\overline{P})$. If M belongs to (P), it will certainly belong to $(\overline{P})$. It follows from what has been said that $(\overline{P})$ must in fact be a closed set. We remark that after inclusion of the points of the boundary (l) in a domain (P), the boundary points can become interior points of the new domain $(\overline{P})$. If, say, (P) is a square with an internal incision, the points of the incision are points of (l), but they become interior points of (P) after including (l) in (P).

We now introduce some concepts referring to any set of points (P) in a plane, and not merely to a domain. We define the *derived set* (P') *of a set* (P) as the set of all limit-points of (P). The proof that *every derived set is closed* follows exactly the same lines as the proof that (l) is closed. Let (P_1) be the set of all the points of the plane not belonging to (P); it is usually called the *complement of* (P). *The boundary* (l) *of a set* (P) is defined as the set of points either belonging to (P) and the derived set of (P_1) or to (P_1) and the derived set of (P), i.e. belonging to (P) and (P_1') or to (P') and (P_1). This definition amounts to the previous one, as regards the boundary of a domain. Another definition of boundary can be given, equivalent to the above. We call M *an isolated point of a set* (P) if there exists an ε-neighbourhood of M which contains no points of (P) apart from M itself. It may easily be seen that the boundary of a set (P) consists of the isolated points of (P) and of the limit-points of (P) that are not interior points of (P). It can be shown, as above, that (l) is a closed set. We shall be dealing in future chiefly with domains.

It should be noted that all the above is applicable to sets of points on a straight line, which can be taken as the axis OX. What we call an ε-neighbourhood of the point $x = c$ now becomes the interval $(c - \varepsilon, c + \varepsilon)$, i.e. the interval of length 2ε with its centre at the given point.

89. Basic theorems regarding sets. A set (P) is said to be *bounded* if all its points lie in a bounded part of the plane. This latter can always be taken as

a square with sides parallel to the axes. We can therefore say that (P) is bounded if all its points belong to such a square.

THEOREM 1. *Every infinite bounded set* (P) *has at least one limit-point.* We prove this first for the case when the points of (P) lie on a straight line, say on the x axis. Set (P) is infinite by hypothesis, i.e. it contains an infinite number of points; also, since (P) is bounded, all its points belong to a finite interval (a, b). We bisect (a, b). At least one half (a_1, b_1) contains an infinite set of points of (P). We now bisect (a_1, b_1). At least one new half (a_2, b_2) contains an infinite set of points of (P), and so on. We get the sequence of intervals

$$(a, b,)\ (a_1, b_1),\ (a_2, b_2), \ldots,\ (a_n, b_n), \ldots,$$

each successive member of which is half the previous one, whilst all the (a_n, b_n) contain an infinite set of points of (P). We know [I, 42] that a_n, b_n have a common limit p. Given any $\varepsilon > 0$, the interval $(p - \varepsilon, p + \varepsilon)$ contains all the (a_n, b_n) as from a certain n, and therefore contains an infinite set of points of (P), i.e. p is a limit-point of (P), which is what we wanted to prove.

We now prove the theorem for a plane. Since (P) is bounded, all its points belong to a square $a < x < b$, $c \leqslant y < d$, which we shall denote symbolically as $[a, b; c, d]$. We divide the square into quarters. At least one quarter $[a_1, b_1; c_1, d_1]$ contains an infinite set of points of (P). We next divide this quarter square into quarters, at least one of which contains an infinite set of points of (P), say $[a_2, b_2; c_2, d_2]$, and so on. We get two sequences of intervals

$$(a, b),\ (a_1, b_1),\ (a_2, b_2), \ldots,\ (a_n, b_n), \ldots$$

$$(c, \mathrm{d}),\ (c_1, \mathrm{d}_1)\ (c_2, \mathrm{d}_2), \ldots,\ (c_n, \mathrm{d}_n), \ldots,$$

in each of which each successive member is half the previous one. Hence, a_n and b_n have a common limit p, whilst c_n, d_n have limit q. By construction, any ε-neighbourhood of the point (p, q) contains all the squares $[a_n, b_n; c_n, d_n]$ as from a certain n, and therefore contains an infinite set of points of (P), i.e. (p, q) is in fact a limit-point of (P).

We take a sequence of ε-neighbourhoods of the point (p, q), with ε taking a decreasing sequence of values $\varepsilon_1, \varepsilon_2, \ldots$ which tend to zero. Let M_1 be a point belonging to (P) in the ε_1-neighbourhood of (p, q). Let M_2 be a point belonging to (P) in the ε_2-neighbourhood of (p, q), differing from M_1. Let M_3 be a point of (P) in the ε_3-neighbourhood, differing from M_1, M_2 and so on. This gives us a sequence of points M_n which will tend to the point $M(p, q)$, i.e. the distance $M_n M$ tends to zero or, what amounts to the same thing, the coordinates (x_n, y_n) of the point M_n tend respectively to the limits p and q. In other words: *a sequence of points, tending to a limit, can be chosen from an infinite bounded set* (P).

We shall only consider bounded sets in future and shall not specially stipulate this. Let (P) and (Q) be two sets, and let us take all the possible distances MN between points M of (P) and points N of (Q). We now have set of non-negative numbers MN, which must have a strict lower bound δ [I, 42]. The non-negative number δ is called *the distance between sets* (P) *and* (Q).

THEOREM II. *If* (P) *and* (Q) *are two closed sets with no common points, the distance* δ *between them is positive.*

We use *reductio ad absurdum*. Let $\delta = 0$. Sets (P) and (Q) have no common points, so that there can be no $MN = 0$. But it follows from the definition of strict lower bound that, for any $\varepsilon > 0$, there exists an M of (P) and an N of (Q) such that $MN < \varepsilon$. We can, therefore, choose a sequence of points M_n of (P) $(n = 1, 2, 3, \ldots)$ and N_n of (Q), such that $M_n N_n \to 0$. Two cases may be imagined as regards the M_n: either there is an infinity of identical points among them, or there is not. In the first case, we retain only the pairs $M_n N_n$ in which the identical M_n appear (there will be more than one set, if there are several such infinite groups of identical points), and we enumerate the pairs by means of integers. In the second case, the infinite bounded set of M_n clearly has a limit-point M and, in accordance with the above, we can choose a sub-sequence of M_n which tends to M. We retain only the pairs $M_n N_n$ in which members of the subsequence appear and enumerate these by means of integers. We carry out a similar process in regard to the N_n. We are now left with pairs of M_n and N_n such that: (1) $M_n N_n \to 0$; (2) M_n tends to M (or coincides with M for all n) and N_n tends to the point N (or coincides with N for all n). We obtain on passing to the limit: $MN = 0$, i.e. M and N coincide. On the other hand, M, as the limit-point of M_n, belonging to (P), is a limit-point of (P) and, since (P) is closed, must belong to (P). Similarly, N must belong to (Q). But M and N coincide, i.e. (P) and (Q) have a common point, which contradicts the original assumption of the theorem. The hypothesis that $\delta = 0$ is therefore false, which proves the theorem.

We carried out the proof for the case when M_n and N_n do not coincide with M and N. If, say, M_n coincides with M for all n, whilst N_n does not coincide with N, we have $MN_n \to 0$, where M belongs to (P). We again have $MN = 0$ in the limit, and the proof remains as before. The case when all the M_n coincide with M and all the N_n with N clearly contradicts the assumption that (P) and (Q) have no common points.

A repetition of the above proof could be used for the following theorem: *if* (P) *and* (Q) *are closed sets, there is at least one pair of points* M *of* (P) *and* N *of* (Q) *such that* $MN = \delta$.

We introduce one further concept. Let us take all the possible distances $M'M''$, where M', M'' are any two points of a given set (P). The set of non-negative numbers $M'M''$ is bounded above, since (P) is bounded, and therefore [I, 42] has a strict upper bound d, which is called the *diameter of the set* (P). *If* (P) *is closed*, it can be shown as above that *at least one pair of points* M', M'' *of* (P) *can be found such that* $M'M'' = d$.

All the above is applicable to three-dimensional space, referred to axes XYZ. An ε-neighbourhood of a point M must be understood here to mean a sphere with centre M and radius ε, whilst we must take the cube $a \leqslant x \leqslant b$, $c \leqslant y \leqslant d$, $e \leqslant z \leqslant f$, instead of the square $a \leqslant x \leqslant b$, $c \leqslant y \leqslant d$.

A square must replace an interval on the x axis.

90. Interior and exterior areas. We take as the basis for the measure of area the statement that the area of a square with sides parallel to the axes is equal to the square of the length of side. We take lines parallel

to the axes, dissecting the plane into a mesh of equal squares. A closed domain consisting of a finite number of squares of the mesh will be called a domain of type (a).

The area of this domain is defined as the sum of the areas of the constituent squares. There is an infinite number of ways of drawing lines parallel to the axes that will dissect a domain of type (a) into corresponding squares. We shall not stop to prove the simple fact that the sum of the areas of these squares is always the same for a given (a) domain. Furthermore, if one or more (a) domains with no common interior points lie inside an (a) domain (A), the sum of the areas of these domains is less than the area of (A). In future, square will be understood to mean the square together with its boundary.

Let (P) be a bounded set of points. On dissecting the plane with a mesh of equal squares, let (S) be the set of all the squares of the mesh, all the points of which (including also points of their boundaries) are interior points of (P).

We use the same letter S to denote the sum of the areas of these squares. Clearly, (S) is a domain of type (a) (Fig. 81). Further, let ($S + S'$) be the set of squares of the mesh having points in common with (P). Let $S + S'$ denote the sum of the areas of these squares.

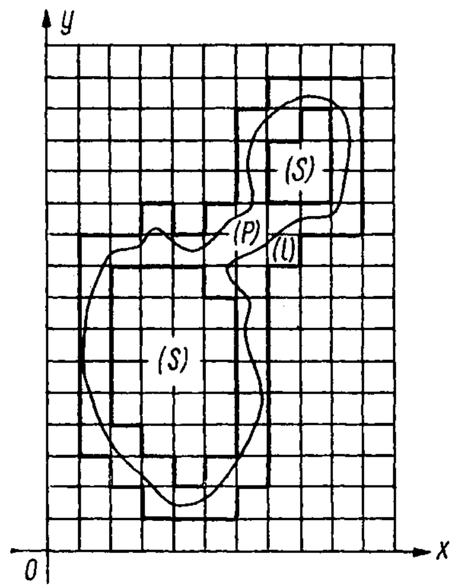

FIG. 81

Evidently, ($S + S'$) has the same structure as (S), and (S) is part of ($S + S'$). The latter set contains, apart from squares appearing in (S), the set (S') of squares having points in common with (l).

On taking all the possible meshes of different squares, we get an infinite set of numbers S. All these numbers will be less than the area of the square, inside which the bounded set (P) is situated. *The upper bound of the set of* S *is called the interior area of set* (P). We denote this by a. Similarly, the set of positive numbers ($S + S'$) has a strict lower bound, which we call *the exterior area of* (P) and denote by A. Finally, let r denote the length of side of a square of the mesh. We prove the following basic theorem [cf. I, 115]:

THEOREM. *If* $r \to 0$, $S \to a$ *and* $S + S' \to A$, i.e. *as the mesh grows indefinitely finer, S tends to the interior area and $S + S'$ to the exterior area.*

We always have $S + S' \geqslant A$ by the definition of strict lower bound. We have to show that, for any given positive ε there exists a positive η such that $S + S' < A + \varepsilon$ if $r < \eta$. From the definition of strict lower bound, a mesh of squares exists such that the corresponding sum $S + S'$, which we denote as $S_0 + S'$, is less than $A + \varepsilon$, i.e. $S_0 + S' < A + \varepsilon$. Let r_0 denote the length of side of a square of this mesh. The boundary of ($S + S'$) consists of a finite number of lines parallel to the axes. We can border ($S_0 + S'$) by squares of

side r_0/n, where n is a positive integer, so as to obtain an (a) domain S_1 formed by squares of side r_0/n which contains $(S_0 + S')$ strictly inside itself. If we make n sufficiently large, the area S_1 of domain (S_1) will differ by as little as required from $S_0 + S'$, so that we can write $S_1 < A + \varepsilon$.

Let (l_1) be the boundary of (S_1) and (l) the boundary of (P). The closed sets (l_1) and (l) have no common points and the distance δ between the sets will be positive. If we take $r < \delta/\sqrt{2}$, all the squares of the mesh having points in common with (P) will clearly lie inside (S_1), and hence, $S + S' < S_1 < A + \varepsilon$ for $r > \delta/\sqrt{2}$. We can therefore take the above η as equal to $\delta/\sqrt{2}$, and we have proved that $S + S' \to A$ as $r \to 0$. The proof that $S \to a$ as $r \to 0$ is exactly similar.

COROLLARY. (S) is part of $(S + S')$, and therefore $S < S + S'$. We get $a \leqslant A$ as $r \to 0$, i.e. *the interior area is not greater than the exterior area.*

If (P) has no interior points, $S = 0$ for any mesh of squares and the interior area is zero. If there are interior points, an ε-neighbourhood of an interior point of (P) will contain a square with sides parallel to the axes, all the points of which are interior points of (P). We shall have $S > 0$ for the corresponding mesh, so that the interior area of (P) will be greater than zero.

If the exterior area of (P) is zero, the interior area will certainly likewise be zero, and it follows that (P) has no interior points in this case. Sets with zero exterior area will be of considerable importance to us later. It follows from the above that these are the sets (P) for which the sum of the areas of the squares of the (closed) meshes having points in common with (P) tends to zero as $r \to 0$.

It can be shown that closed curves exist which do not cut themselves and which have the parametric equations $x = \varphi(t)$, $y = \psi(t)$, where $\varphi(t)$ and $\psi(t)$ are continuous functions, whilst their exterior areas are greater than zero. These curves are the boundaries of connected domains, as may be shown, and interior areas are less than exterior areas in such domains.

91. Measurable sets. *A set (P) is said to be measurable if $a = A$,* i.e. *if its interior and exterior areas are equal.* The common value of a and A is called the *area of set* (P). We notice that, if the exterior area is zero ($A = 0$), $a = 0$, as we saw above, i.e. the set here is measurable and its area is zero. Conversely, if a set is measurable and its area is zero, it is obvious that $A = 0$.

The necessary and sufficient condition for a set to be measurable is evidently that S and $S + S'$ have the same limit as $r \to 0$, in other words, that $S' \to 0$ as $r \to 0$. This means that, given any positive ε, there exists a positive η such that $S' < \varepsilon$ if $r < \eta$.

If the boundary (l) of (P) has zero area, $S' \to 0$ as $r \to 0$, since all the (closed) constituent squares of S' have points in common with (l). The converse is also true, i.e. if $S' \to 0$ as $r \to 0$, the area of (l) is zero. The proof is as follows. Squares of any mesh that have points in common with (l) can only be absent from (S') in the case when their only common points with (l) are on their boundaries, since if points of (l) lay inside a square, points of (P) would also lie inside, i.e. the square would belong to (S'). Squares of this sort — having points in common with (l) only on their boundaries, and neither containing points of (P) nor belonging to (S') — can in fact exist (there is a finite number of such

squares), though there must be points of (P) in at least one of the eight squares that border such a square. If we add to all the squares that belong to (S') all the squares that border on them, we must obtain all the squares without exception having points in common with (l). The area of all the squares thus obtained is not greater than $9S'$, and, since $S' \to 0$ by hypothesis, as $r \to 0$, we can say that the exterior area of (l) is zero, i.e. simply, the area of (l) is zero. We get the following important theorem from the above.

THEOREM. *A necessary and sufficient condition for a set (P) to be measurable is that the area of its boundary (l) is zero,* i.e. *it is necessary and sufficient for the sum of the areas of the squares of the mesh that have points in common with (l) to tend to zero as $r \to 0$.*

Remark. This theorem can be re-stated as follows by using the theorems of [90]: *the necessary and sufficient condition for (P) to be measurable is that, for any given positive ε, there exists a mesh of squares such that the sum of areas of squares of the mesh with points in common with (l) is less than ε.*

If the exterior area $A = 0$ for a given set, we have all the more $A = 0$ for part of the set, i.e. any part of a set with zero area also has zero area.

We note a further property of sets of zero area. By using the above method of surrounding a square with its neighbouring squares, it can easily be shown that, for any given positive ε, a set with zero area can strictly enclose an (α) domain of area less than ε.

We now consider domains, instead of sets in general.

Let a measurable domain (P) be divided into two parts (P_1) and (P_2) with the aid of a set (of lines) (λ), the exterior area of which is zero. This means that interior points of (P_1) and (P_2) are interior points of (P), not belonging to (λ).

It follows from the above that (P_1) and (P_2) are measurable, whilst the sum of their areas is equal to the area of (P). The same is true when (P) is divided into any finite number of domains. Conversely, if we combine into one set any finite number of measurable closed domains (or sets) (P_k) with no common interior points, the new set (P) is measurable, and its area is equal to the sum of the areas of the original sets. Points on the boundaries of the (P_k) may become interior points of the total set. If a measurable domain (set) (Q_1) is part of a measurable domain (set) (Q_2), i.e. if every point of (Q_1) belong to (Q_2), the area of (Q_1) cannot exceed that of (Q_2). All this follows from the above definitions and theorems.

We shall in future take it for granted that the sub-division of a measurable set is carried out with the aid of a set of points with zero exterior area.

A simple example may be given of a curve (λ) with zero exterior area, i.e. such that the sum of the areas of squares of the mesh having points in common with (λ) tends to zero along with r; at the same time, we take (λ) to have the explicit equation $y = \varphi(x)$, where x varies in a finite interval (a, b) and $\varphi(x)$ is continuous in this interval. By uniform continuity, for any given positive ε there exists a δ such that $|\varphi(x'') - \varphi(x')| < \varepsilon/3(b-a)$ if x', x'' belong to (a, b) and $|x'' - x'| < \delta$ [I, 43]. We choose r less than δ and less than $\varepsilon/3(b-a)$. We construct the mesh by dissecting (a, b) with the points $a = x_0 < x_1 < < x_2 < \ldots < x_{n-1} < x_n = b$, where the lengths $(x_k - x_{k-1})$ are all equal to r except for the extreme dissections, $(x_1 - a)$ and $(b - x_{n-1})$ being possibly less

than r (Fig. 82). We consider the squares of the mesh lying in the band between $x = x_{k-1}$ and $x = x_k$. Since $x_k - x_{k-1} < \delta$, it can be asserted that the difference ω_k between the greatest and least values of $\varphi(x)$ in the interval (x_{k-1}, x_k) (i.e. the oscillation of $\varphi(x)$ in the interval) is less than $\varepsilon/3(b - a)$. The square with a point in common with the lowest point of $y = \varphi(x)$ can go at the most a distance r lower (the length of side of the square), whilst the square with a point in common with the highest point of the curve can go higher by r at the most. The sum of the heights of the squares that have points in common with (λ) and that lie in the band $x = x_{k-1}, x = x_k$, is thus less than $\varepsilon/3(b - a) + 2r$, or less than $\varepsilon/(b - a)$, since $r < \varepsilon/3(b - a)$, whilst the sum of their areas is less than $\varepsilon(x_k - x_{k-1})/(b - a)$. We see, on summing from $k = 1$ to $k = n$, that the sum of the areas of the squares with points in common with (λ) is less than ε, and it follows, since ε is arbitrary, that the exterior area of (λ) is zero. The proof is the same, that a curve with explicit equation $x = \psi(y)$, where $\psi(y)$ is continuous in a finite interval, also has zero exterior area. A *simple curve* is defined as one which has an equation $y = \varphi(x)$ or $x = \psi(y)$, where $\varphi(x)$ or $\psi(y)$ is continuous in the corresponding finite interval of variation of the independent variable.

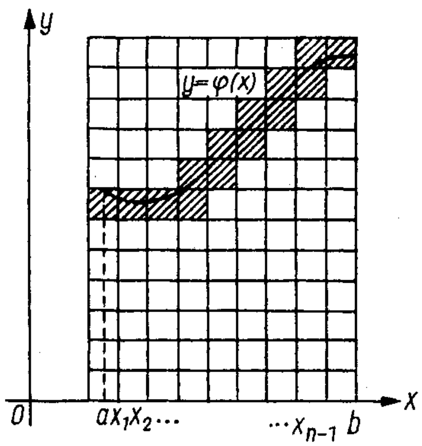

FIG. 82

It follows from the above that *the exterior area of a simple curve is zero.* Hence follows a sufficient test as to whether a domain is measurable.

THEOREM. *If the boundary of a domain (or set) (P) is a simple curve, (P) is measurable.*

It follows from the above discussion that, if we divide a measurable domain into a finite number of domains by means of a simple curve (or, what amounts to the same thing, a finite number of simple curves), each new domain is also measurable, and the sum of the areas of these is equal to the area of the original domain. It must be remembered that what has been said is applicable to the division of any measurable set into a finite number of measurable parts.

It is easily shown that the definite integral $\int_a^b \varphi(x)\,dx$ gives the area, in the above sense of the word, of the domain bounded by the curve $y = \varphi(x)$, the x axis, and the straight lines $x = a$, $x = b$, it being assumed that $\varphi(x) > 0$.

92. Independence on the choice of axes. The definition of (interior and exterior) area is closely bound up with the choice of axes, inasmuch as our measurements were made with a mesh of squares with sides parallel to the axes. Parallel displacement of the axes evidently plays no part in the measurements, but rotation of the axes about the origin makes a fundamental difference, since (P) now has to be covered with a different mesh of squares. Instead of turning the axes counter-clockwise by an angle φ, we can keep them fixed and rotate (P) by an angle $(-\varphi)$ about the origin. It is evident from this that the independence of the area on the choice of axes amounts to the area being unchanged on

displacement of (P) as a whole over the plane. The case of parallel displacements is obvious, but that of rotation about the origin requires proof.

We can prove the following theorem, analogous to that of [90].

THEOREM I. *Let the complete plane be divided into measurable domains* (Δ_ν) *with diameters not greater than a number* d *and such that any bounded part of the plane has points in common with only a finite number of these domains. Let* Σ *denote the sum of the areas of those domains, every point of which, including the boundary points, is an interior point of* (P)*; also, let* $\Sigma + \Sigma'$ *denote the sum of the areas of domains with at least one point in common with* (P)*. Then if* $d \to 0$, Σ *tends to the interior area of* (P)*, whilst* $\Sigma + \Sigma'$ *tends to the exterior area of* (P).

The implication of this theorem is that, in evaluating the exterior and interior areas, we can use any mesh of measurable domains, provided $d \to 0$, instead of a mesh of squares with sides parallel to the axes.

We note that the boundary of a square, however disposed relatively to the axes, is a simple curve, and any square is therefore a measurable domain. The above theorem, which we shall only use in the case when the squares of our mesh are equal, has the direct consequence that the area of a domain may be measured by using say a mesh of squares with sides not parallel to the axes, provided the side tends to zero. But here we need to know the value of the area of a square with sides not parallel to the axes; that this area is equal to the square of the length of side is still not explicit, strictly speaking, since our basic theorem on the measurement of area assumed a square with sides parallel to the axes. If we can prove that the area of any square is equal to the square of the side, we can assert, by the above, that measurability and area are both independent of the choice of direction of the axes, and that an area is unaffected by a rotation.

All this leads us to proving Theorem II.

THEOREM II. *The area of a square with sides parallel to the axes is unchanged on rotation about the origin.* We first of all recall that the boundary of any square is a simple curve, so that any square is a measurable domain. Let (q) be the initial square, with side r, and let (q_1) be the square obtained by rotation. We use the letters without brackets to denote the corresponding areas and let $q_1/q = s$. The ratio q_1/q will be the same for a given rotation of the axes for all squares of side r, since we can make any parallel square of side r coincide with (q) by a parallel shift of the axes, which does not affect area. We now carry out a transformation of similitude on the plane, with centre at the origin, in which the length of any radius vector from the origin is multiplied by some positive number k. The transformation amounts to shifting a point with coordinates (x, y) to the point (kx, ky) [3]. All dimensions of length will be multiplied by k. Every square with sides parallel to the axes becomes a similar square with k times the length of side. It follows that an (interior or exterior) area is multiplied by k^2. Let (q'), (q_1') denote the squares obtained by the transformation of similitude from (q), (q_1) respectively. Obviously, (q_1') is found from (q') by the same rotation as that giving (q_1) from (q). But $q_1' = k^2 q_1$ and $q' = k^2 q$, so that $q_1'/q' = s$. Now we can transform the square (q) into a square with any length of side by a suitable choice of k. Hence we see that the ratio $q_1/q = s$ is the same for all initial squares (q) for a given rotation of the plane. We now prove that $s = 1$. We take the circle $x^2 + y^2 < 1$ with centre at the

origin and unit radius, covered with a mesh of squares with sides parallel to the axes. This circle is clearly a measurable domain.

The area of a square is multiplied by s on rotation about the origin, and by the definition of area and the theorem proved above, the area of the circle must also be multiplied by s. But the circle is unaffected by rotation, so that its area must be unchanged, i.e. $s = 1$, which proves our theorem.

93. The case of any number of dimensions. The entire theory of areas can be carried over to the case of three-dimensional space, giving us interior and exterior volumes and measurable three-dimensional domains or sets. Cubes perform the role of squares.

We can construct an analogous theory of the measurement of "area", or theory of measure, for any n-dimensional space. A point becomes the set of n real numbers $(x_1, x_2, \ldots, x_n)$, arranged in a definite order. The distance between two points $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ is given by the expression:

$$r = \sqrt{\sum_{s=1}^{n} (y_s - x_s)^2}.$$

A sphere with centre $(a_1, a_2, \ldots, a_n)$ and radius ϱ is defined as the system of points $(x_1, x_2, \ldots, x_n)$ whose coordinates satisfy the inequality

$$\sum_{s=1}^{n} (x_s - a_s)^2 < \varrho^2.$$

Finally, a cube with edge r is defined as the system of points whose coordinates satisfy the inequalities $a_s < x_s < b_s$ $(s = 1, 2, \ldots, n)$, where $b_s - a_s = r$. We call r^n the measure of the cube. These definitions together enable us to re-state all the above theory for n-dimensional space and establish the concepts of interior and exterior measure of a domain in general, as also to speak of a domain (or set) being measurable. All the theorems proved are valid for n-dimensional space. A parallel shift is now given by $x'_s = x_s + a_s$ $(s = 1, 2, \ldots, n)$, whilst rotation about the origin is expressed by a linear transformation, in which the distances of points from the origin remain unchanged. A more detailed treatment of these transformations will be found in Volume III.

We used the idea of a step-line, i.e. a line composed of a finite number of pieces of straight lines, in defining a connected domain. A straight line in n-dimensional space is defined as a line (i.e. a set of points) having the parametric equations $x_s = \varphi_s(t)$, where the $\varphi_s(t)$ are first degree polynomials. A domain in n-dimensional space is exemplified by the set of interior points of a sphere or a cube. Such a domain is usually defined by certain inequalities which the coordinates of its points must satisfy. We remark that, with $n = 1$, i.e. on a straight line, a connected domain must mean the set of interior points of a certain interval. What we have said about simple curves is easily generalized for n dimensions. In particular, if $z = \varphi(x, y)$ is the explicit equation of a given surface in three-dimensional space, where $\varphi(x, y)$ is continuous in some bounded closed domain of the XY plane, the surface is a measurable set and its measure is zero. The further concept of a simple surface is easily built up after the manner of [91], and any domain bounded by a simple surface will be measurable.

94. Darboux's theorem. Having established the concept of area, we turn to the theory of double integrals. All our discussion will equally apply to triple integrals, and will be arranged along the same lines as used for ordinary integrals in Volume I; detailed arguments will be omitted when precisely analogous to those of Volume I.

Let (σ) be a measurable domain in the plane and let $f(N)$ be a bounded function, defined at all points of the closed domain (σ). We divide (σ) into a finite number of measurable domains (σ_k) $(k = 1, 2, \ldots, n)$ and let σ and σ_k denote as usual the corresponding areas, so that $\sigma = \sigma_1 + \sigma_2 + \ldots + \sigma_n$. We take the diameter of any (σ_k) to be less than some number d and let N_k be any point belonging to the closed domain (σ_k).

We form the sum of products:

$$\sum_{k=1}^{n} f(N_k)\,\sigma_k. \tag{1}$$

The type of function $f(N)$ for which this sum has a definite limit as $d \to 0$ will be discussed later. Let M_k and m_k be the strict upper and lower bounds of $f(N)$ in the closed (σ_k). As well as the sum (1), we form:

$$S = \sum_{k=1}^{n} M_k\,\sigma_k \tag{2}$$

$$s = \sum_{k=1}^{n} m_k\,\sigma_k. \tag{3}$$

We have, as in [I, 115]:

$$s < \sum_{k=1}^{n} f(N_k)\,\sigma_k < S, \tag{4}$$

and we can say that, for any dissection of (σ), S and s lie between the bounds $m\sigma$ and $M\sigma$, where M and m are the strict upper and lower bounds of $f(N)$ in the closed (σ).

We now consider S in more detail, and assume that $f(N)$ is always positive. Let three measurable domains $(\sigma_k^{(1)})$, $(\sigma_k^{(2)})$, $(\sigma_k^{(3)})$ with no common interior points lie inside (σ_k), and let the strict upper bounds of $f(N)$ in the new domains be respectively $M_k^{(1)}, M_k^{(2)}, M_k^{(3)}$. Since $M_k^{(1)}, M_k^{(2)}, M_k^{(3)} < M_k$, $\sigma_k^{(1)} + \sigma_k^{(2)} + \sigma_k^{(3)} < \sigma_k$, and all $f(N)$ are positive, we can write:

$$M_k^{(1)}\sigma_k^{(1)} + M_k^{(2)}\sigma_k^{(2)} + M_k^{(3)}\sigma_k^{(3)} \leqslant M_k\,\sigma_k. \tag{5}$$

Let L be the strict lower bound of all possible values of S. We prove that $S \to L$ as $d \to 0$. It is sufficient to show that, for any positive ε, there exists a positive η such that $S < L + \varepsilon$ provided $d < \eta$. By definition of L, there exists a fully defined rule (I) for dissection of (σ) into (σ_k') such that the value of S thus obtained, say S', is less than $L + \varepsilon/2$. Let (λ_0) be the closed set of points consisting of points of the boundary of (σ) and of the boundaries of all the (σ_k'). The area of (λ_0) is zero by the definition of measurability, and we can enclose (λ_0) strictly inside a finite number of squares, the sum of the areas of which is less that $\varepsilon/2M$. Let (l_0) be the boundary of the closed domain (Q_0)

formed by these squares. Let δ be the positive distance between the closed sets (λ_0) and (l_0), which have no common points. We show that it is sufficient to take $\eta = \delta/2$. Let us take $d < \delta/2$ for the dissection of (σ). We divide the sub-domains obtained into two classes: those that have no common points with (λ_0) go into the first class, and the remainder go into the second. Let us call sub-domains of the first class (σ_s), and of the second class (τ_m). The sum S is now divided into two sums: $S = S_1 + S_2$, where

$$S_1 = \sum' \mu_l \sigma_l; \quad S_2 = \sum' \nu_m \tau_m,$$

and μ_l and ν_m are the strict upper bounds of $f(N)$ in the closed (σ_l) and (τ_m) respectively.

Every (σ_l) lies inside a certain (σ'_k) of the first rule (I) for dissection of (σ), and by (5), the sum of the terms of S_1 for which (σ_l) lies inside (σ'_k) is not greater than $M_k \sigma'_k$, so that $S_1 \leqslant S'$, i.e. since $S' < L + \varepsilon/2$, we have $S_1 < L + \varepsilon/2$. As regards S_2, the (τ_m) have points in common with (λ_0) and their diameters are less than $\delta/2$; hence, they all lie inside the squares forming (Q_0). It follows that $\Sigma\tau_m$ cannot exceed the sum of the areas of these squares, i.e. $\Sigma\tau_m \leqslant \varepsilon/2M$. We have $\nu_m \leqslant M$, so that

$$S_2 = \sum \nu_m \tau_m \leqslant M \frac{\varepsilon}{2M} = \frac{\varepsilon}{2}.$$

The required inequality $S < L + \varepsilon$ for $d < \eta$ follows from $S_1 < L + \varepsilon/2$ and $S_2 \leqslant \varepsilon/2$. Our assertion that $S \to L$ is proved for positive functions. We see that it is true for any bounded function by repeating the argument of [I, 115]. Similarly, we can show that $s \to l$ as $d \to 0$, where l is the strict upper bound of s. This gives us the following theorem.

DARBOUX'S THEOREM. *On indefinite decrease of the greatest of the diameters of the sub-domains (σ_k), the sums s and S tend to definite limits l and L, where $l \leqslant L$.*

The whole of the above discussion is applicable word for word to the case when (σ) and (σ_k) are any measurable sets. Darboux's theorem remains valid.

95. Integrable functions. We call $f(N)$ an integrable function over (σ) if the sum

$$\sum_{k=1}^{n} f(N_k)\, \sigma_k \tag{6}$$

has a definite limit as the greatest of the diameters d of the sub-domains (sub-sets) (σ_k) tends to zero. The limit is called the double integral of $f(N)$ over the domain (set) (σ):

$$\iint\limits_{(\sigma)} f(N)\, d\sigma = \lim \sum_{k=1}^{n}{}' f(N_k)\, \sigma_k.$$

It can be shown, as in [I, 116], that *the necessary and sufficient condition for $f(N)$ to be integrable is that the limits l and L of the sums s and S coincide*, i.e. *that the difference between the sums:*

$$\sum_{k=1}^{n}{}' (M_k - m_k)\, \sigma_k \tag{7}$$

tends to zero as $d \to 0$. The common limit of s and S is the value of the integral.

If $f(N) = 1$, (6) is always equal to the area σ of the domain (set) (σ), i.e.

$$\iint_{(\sigma)} d\sigma = \sigma.$$

We also note that, by Darboux's theorem, the above condition for integrability can be re-stated as: *for any given positive ε there exists a dissection of (σ) into sub-domains such that expression* (7) *is less than ε.* Use of the condition for integrability reveals certain classes of integrable functions.

1. *If $f(N)$ is continuous in a closed domain (set) (σ), it is integrable.* The proof is the same as in [I, 116].

2. Now let $f(N)$ have points of discontinuity, whilst being bounded as before, in the closed domain (σ). If we assume that the area of the set (R_0) of points of discontinuity is zero, we can show that $f(N)$ is integrable.

We first let (σ) be a square with sides parallel to the axes and we let it be dissected into squares; the case of any measurable domain will be considered in the next section. Let ε be a given positive number; (σ) can be divided into equal squares in such a way that every point of (R_0) lies strictly inside an (a) type domain, the area of which is less than $\varepsilon/2\mu$, where $\mu = M - m$ [cf. 91]. We have uniform continuity of $f(N)$ in the remaining closed (a) domain, since this contains no points of (R_0). Let these (a) domains be (E) and (F). Since $f(N)$ is uniformly continuous in the closed domain $(\overline{F})$, this can be dissected into as small sub-domains as required, such that the sum of the terms of (7) which relate to these is less than $\varepsilon/2$. On taking into account that $M_k - m_k \leqslant M - m = \mu$, and that the area of $(E) < \varepsilon/2\mu$, we can see that the terms of (7) relating to (E), which cannot be sub-divided, is less than $\varepsilon/2$; hence, the whole of (7) is less than ε, so that $f(N)$ is integrable.

It follows that, *if the set of points of discontinuity of a bounded function $f(N)$ has zero area, $f(N)$ is integrable. This condition is certainly satisfied if $f(N)$ has a finite number of points of discontinuity or if the points of discontinuity lie on a finite number of simple curves.*

96. Properties of integrable functions. We briefly indicate the properties of integrable functions, corresponding to [I, 117] for single integrals.

I. *If $f(N)$ is integrable in a measurable domain (σ) and we change the value of $f(N)$ at a set of points (R_0) of zero area whilst preserving the boundedness of the function, the new function is also integrable and the value of the integral remains unchanged.*

The proof is much the same as at the end of the previous article when (σ) is a square with sides parallel to the axes. The (a) domains (E) and (F) are distinguished as before. The value of $f(N)$ is unchanged in (F), and since $f(N)$ is integrable, the terms of (7) relating to (F) have a sum less than $\varepsilon/2$ for a sufficiently fine dissection of (F). The terms relating to (E) are less than $\varepsilon/2$ by the smallness of the area of (E), and hence follows the integrability of the new function, as above. Further, the set (R_0) has no interior points, and there is a point N_k in each of the (σ_k) acquired by dissection of (σ) at which the value of $f(N)$ is unchanged. We see by using these points in forming the sum (6) that the value of the integral is unchanged.

II. *If $f(N)$ is integrable in a measurable domain (σ) and the domain is divided into a finite number of measurable domains (σ_1), (σ_2), ..., (σ_n), $f(N)$ will be integrable in each of the (σ_k) and the integral over (σ) is equal to the sum of the integrals over the (σ_k).*

We dissect (σ) by subdividing the (σ_k). The sum (7), consisting of non-negative terms, will tend to zero by the integrability of $f(N)$. The sum of the terms relating to each particular (σ_k) will all the more tend to zero, i.e. $f(N)$ is integrable over (σ_k). The second part of the statement follows directly, if passage to the limit in sum (6) is carried out for each summation relating to each particular (σ_k). The converse is obviously also true: *integrability over the (σ_k) implies integrability over (σ)*. The other properties of integrals given in [I, 117] also remain valid; these relate to the removal of a constant factor outside the integral sign, to the integrability of the sum, product and quotient of integrable functions, as also of the absolute value of an integrable function. The proof of the mean value theorem follows the same lines as in [I, 95].

We next turn to the proof of the second condition for integrability [95] and of the first property of the present section for the case of any measurable domain (σ). Let the bounded function $f(N)$ be given in this domain and on its boundary, the set (R_0) of points of discontinuity of the function having zero area. We draw a square (Q) with sides parallel to the axes and with (σ) lying strictly inside, and we take a new function $f_1(N)$, defined as follows in (Q): $f_1(N) = f(N)$ in the closed (σ) and $f_1(N) = 0$ elsewhere. The points of (Q) belonging neither to its boundary nor to the closed (σ) can easily be seen to form an open set (domain), say (σ_1). The boundary points of (σ_1) belong to the boundary (l) of (σ) and to the boundary of (Q). Since (σ) is measurable, the area of (l) is zero and the same can be said of the area of the boundary of (Q).

Hence the boundary of (σ_1) also has zero area, i.e. (σ_1) is measurable. The points of discontinuity of $f_1(N)$ in (Q) are the points of (R_0) and also, possibly, points on (l). The area of the set of points of discontinuity of $f_1(N)$ in (Q) is zero in either case, so that $f_1(N)$ is integrable in the square (Q) [95], in other words, it is integrable over (σ) and (σ_1).

But $f_1(N) = f(N)$ in the closed (σ), so that $f(N)$ is integrable over (σ), which it was required to prove. The proof of the first property of integrable functions, carried out above for a square, may be carried out on similar lines for $f(N)$ and any measurable (σ).

We return to $f_1(N)$, which is zero at interior points of (σ_1), so that its integral over (σ_1) is zero, whence:

$$\iint_{(\sigma)} f(N)\,\mathrm{d}\sigma = \iint_{(Q)} f_1(N)\,\mathrm{d}\sigma.$$

It may be remarked that, since the boundary (l) of the measurable (σ) has zero area, the value of the bounded $f(N)$ on (l) does not affect the value of the integral.

97. Evaluation of double integrals. We now establish the formula for reducing a double integral to two quadratures. We first take the case of a rectangle (R) with sides:

$$x = a; \quad x = b; \quad y = c; \quad y = d, \tag{8}$$

parallel to the axes. We let $f(N) = f(x, y)$ be integrable over (R), i.e. there exists

$$\iint_{(R)} f(N)\, \mathrm{d}\sigma = \iint_{(R)} f(x, y)\, \mathrm{d}x\, \mathrm{d}y. \tag{9}$$

We further suppose the existence, for every x of the interval (a, b), of

$$F(x) = \int_c^d f(x, y)\, \mathrm{d}y, \quad (a \leqslant x \leqslant b) \tag{10}$$

and of the iterated integral:

$$\int_a^b F(x)\, \mathrm{d}x = \int_a^b \Big[\int_c^d f(x, y)\, \mathrm{d}y \Big]\, \mathrm{d}x. \tag{11}$$

We dissect (R) by sub-dividing the intervals with the points

$$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$$
$$c = y_0 < y_1 < y_2 < \ldots < y_{m-1} < y_m = d$$

and we let (R_{ik}) denote the sub-rectangle of the dissection, bounded by: $x = x_i$, $x = x_{i+1}$, $y = y_k$, $y = y_{k+1}$. Also, let m_{ik} and M_{ik} denote the strict lower and upper bounds of $f(x, y)$ in the closed (R_{ik}), and let $\Delta x_i = x_{i+1} - x_i$, $\Delta y_k = y_{k+1} - y_k$. On integrating the inequality

$$m_{ik} \leqslant f(x, y) \leqslant M_{ik} \quad [(x, y) \quad \text{from} \quad (R_{ik})]$$

over the interval $y_k \leqslant y \leqslant y_{k+1}$, we get:

$$m_{ik} \Delta y_k \leqslant \int_{y_k}^{y_{k+1}} f(x, y)\, \mathrm{d}y \leqslant M_{ik} \Delta y_k \quad (x_i \leqslant x \leqslant x_{i+1}),$$

where (y_k, y_{k+1}) is part of (c, d) and the integral exists since (10) exists [I, 117]. We obtain on adding these inequalities:

$$\sum_{k=0}^{m-1} m_{ik} \Delta y_k \leqslant \int_c^d f(x, y)\, \mathrm{d}y \leqslant \sum_{k=0}^{m-1} M_{ik} \Delta y_k.$$

We integrate over the interval (x_i, x_{i+1}):

$$\sum_{k=0}^{m-1} m_{ik} \Delta y_k \Delta x_i \leqslant \int_{x_i}^{x_{i+1}} \Big[\int_c^d f(x, y)\, \mathrm{d}y\Big]\, \mathrm{d}x \leqslant \sum_{k=0}^{m} M_{ik} \Delta y_k \Delta x_i;$$

the integral just written exists, since (11) exists. We sum these latter inequalities over i:

$$\sum_{i=0}^{n-1} \sum_{k=0}^{m-1} m_{ik} \Delta y_k \Delta x_i \leqslant \int_a^b \Big[\int_c^d f(x, y)\, \mathrm{d}y\Big]\, \mathrm{d}x \leqslant \sum_{i=0}^{n-1} \sum_{k=0}^{m-1} M_{ik} \Delta y_k \Delta x_i.$$

On noticing that $\Delta y_k\, \Delta x_i$ is the area of (R_{ik}), we can say that the extreme terms of the inequality tend to integral (9) as the sub-rectangles diminish indefinitely, which proves the required formula:

$$\iint_{(R)} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = \int_a^b \Big[\int_c^d f(x, y)\, \mathrm{d}y\Big]\, \mathrm{d}x, \tag{12}$$

i.e. *if double integral* (9) *and iterated integral* (11) *exist,* (12) *is valid and the integrals are equal.*

We remark that the existence of integral (11) presupposes the existence of (10). If $f(N)$ is continuous in the closed rectangle (R), (9) and (10) clearly exist ([95] and [I, 116]). With this, as we saw in [80], (10) is a continuous function of x, so that (11) also exists. We now consider a domain (σ), bounded by the curves $y = \varphi_2(x)$, $y = \varphi_1(x)$, and by the lines $x = a$, $x = b$ (Fig. 83). We assume the existence of the double integral

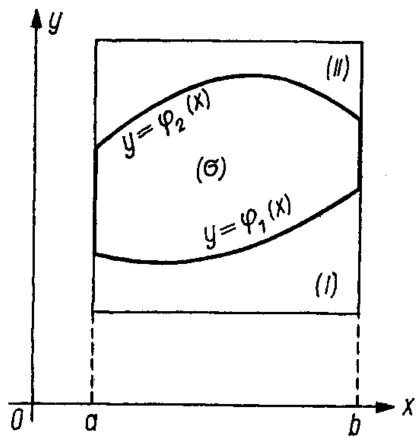

FIG. 83

$$\iint\limits_{(\sigma)} f(N)\, d\sigma = \iint\limits_{(\sigma)} f(x, y)\, dx\, dy, \tag{13}$$

of the single integral

$$F(x) = \int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\, dy \tag{14}$$

and of the iterated integral

$$\int\limits_a^b F(x)\, dx = \int\limits_a^b \Big[\int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\, dy \Big]\, dx. \tag{15}$$

Let (R) be the rectangle formed by the lines (8), where c, d are chosen such that $c > \varphi_1(x)$, $d > \varphi_2(x)$ for all x of (a, b), i.e. (σ) forms part of (R). We define the function $f_1(N) = f_1(x, y)$ in (R), equal to $f(N)$ at points of (σ) and zero throughout the remainder of (R). The curves $y = \varphi_2(x)$, $y = \varphi_1(x)$ divide (R) into three parts: (σ), and domains (I) and (II), lying below and above (σ) (Fig. 83). We have $f_1(N)$ integrable in (σ), since it coincides there with $f(N)$, and integrable in (I) and (II), since it is zero at interior points of these domains.

Hence $f_1(N)$ is integrable in (R) [96] and

$$\iint\limits_{(R)} f_1(N)\, d\sigma = \iint\limits_{(\sigma)} f(N)\, d\sigma. \tag{16}$$

Similarly, for every x of (a, b), we have the existence of

$$F(x) = \int\limits_c^d f_1(x, y)\, dy = \int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\, dy \tag{17}$$

and of integral (15).

It follows that (12) is valid for $f_1(N)$, and by (16) and (17), the reduction of a double integral over (σ) to an iterated integral is obtained from (12) as:

$$\iint\limits_{(\sigma)} f(x, y)\, dx\, dy = \int\limits_a^b \Big[\int\limits_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\, dy \Big]\, dx. \tag{18}$$

We assumed the *existence of integrals* (13), (14) and (15) for this argument. If $f(x, y)$ is integrable in the closed domain (σ), integrals (13) and (14) exist, as above. Moreover, by [80], (14) defines a continuous function in x, and therefore (15) also exists. The formula for reducing a triple to an iterated integral, the latter consisting of three quadratures [58], can be proved in the same sort of way.

98. n-tuple integrals. All the remarks of [94] and [95] can be carried over directly to the case of n dimensional space, leading to the concept of the integral of a bounded function over a bounded measurable domain in n dimensions, as well as to the condition for integrability given above and to the usual properties of integrals. A formula exists, similar to that of [97], for reducing an n-tuple integral to an iterated integral, made up of n quadratures. The formula can be proved by induction, n being varied by unity. The limits of the multiple integral are obtained from the inequalities that define the domain of integration. Let $f(N) = f(x_1, x_2, \ldots, x_n)$ be continuous in a closed measurable domain (P_n) in n dimensions, interior points of the domain being defined by the conditions: The points $(x_1, x_2, \ldots, x_{n-1})$ are interior points of a certain measurable domain Q_{n-1} of $(n-1)$-dimensional space, and the x_n satisfy the inequalities

$$\varphi_1(x_1, x_2, \ldots, x_{n-1}) < x_n < \varphi_2(x_1, x_2, \ldots, x_{n-1}),$$

where $\varphi_1(x_1, x_2, \ldots, x_{n-1})$ and $\varphi_2(x_1, x_2, \ldots, x_{n-1})$ are continuous functions in Q_{n-1}. The n-tuple integral is now given by a quadrature with respect to x_n and an $(n-1)$-tuple integral over (Q_{n-1}):

$$\iint_{(P_n)}\ldots\int f(x_1, x_2, \ldots, x_n)\,dx_1\ldots dx_n =$$

$$= \iint_{(Q_{n-1})}\ldots\int \Big[\int_{\varphi_1(x_1,\ldots,x_{n-1})}^{\varphi_2(x_1,\ldots,x_{n-1})} f(x_1, \ldots, x_n)\,dx_n\Big]\,dx_1\ldots dx_{n-1}. \tag{19}$$

The generalization to n dimensions of a plane rectangle with sides parallel to the axes is the prismatoid (R_n), defined by:

$$a_1 \leqslant x_1 \leqslant b_1; \quad a_2 \leqslant x_2 \leqslant b_2; \ldots; \quad a_n \leqslant x_n \leqslant b_n. \tag{20}$$

Integration over this prismatoid reduces to an iterated integral in which all the limits are constant:

$$\iint_{(R_n)}\ldots\int f(x_1, \ldots, x_n)\,dx_1\ldots dx_n = \int_{a_1}^{b_1} dx_1 \ldots \int_{a_{n-1}}^{b_{n-1}} dx_{n-1} \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\,dx_n,$$

where the order of the integrations can be varied arbitrarily, provided the limits remain unchanged.

The formula for change of variables in an n-tuple integral may be mentioned for the benefit of readers familiar with determinants. Let the variables $(x_1, x_2, \ldots, x_n)$ be replaced by new variables $(x_1', x_2', \ldots, x_n')$, where

$$x_i = \varphi_i(x_1', x_2', \ldots, x_n') \quad (i = 1, 2, \ldots, n). \tag{21}$$

We now bring in the functional determinant of system (21):

$$D = \begin{vmatrix} \dfrac{\partial\varphi_1}{\partial x_1'}, & \dfrac{\partial\varphi_1}{\partial x_2'}, & \ldots, & \dfrac{\partial\varphi_1}{\partial x_n'} \\ \dfrac{\partial\varphi_2}{\partial x_1'}, & \dfrac{\partial\varphi_2}{\partial x_2'}, & \ldots, & \dfrac{\partial\varphi_2}{\partial x_n'} \\ \cdot & \cdot & \cdot & \cdot \\ \dfrac{\partial\varphi_n}{\partial x_1'}, & \dfrac{\partial\varphi_n}{\partial x_2'}, & \ldots, & \dfrac{\partial\varphi_n}{\partial x_n'} \end{vmatrix}. \tag{22}$$

The formula for change of variables runs

$$\underset{(P_n)}{\int\int\ldots\int} f\mathrm{d}x_1\ldots\mathrm{d}x_n = \underset{(P'_n)}{\int\int\ldots\int} f|D|\,\mathrm{d}x'_1\ldots\mathrm{d}x'_n, \tag{23}$$

where the inequalities defining the new domain of integration (P'_n) are obtained from those defining (P_n) by replacing the x_i appearing there by expressions (21). The conditions for (23) to be applicable are the same as laid down for double integrals in [77]. Improper n-tuple integrals are defined in the same way as improper double and triple integrals [86]. We now turn to some examples.

99. Examples. 1. A tetrahedron in n-dimensional space, bounded by the hyperplanes

$$x_1 = 0;\ x_2 = 0;\ldots;\ x_n = 0;\ x_1 + x_2 + \ldots + x_n = a\ (a > 0),$$

s defined by the inequalities:

$$x_1 > 0;\ x_2 > 0;\ldots;\ x_n > 0;\ x_1 + x_2 + \ldots + x_n < a. \tag{24}$$

With $n = 3$, an ordinary tetrahedron is obtained, bounded by the coordinate planes and by the plane $x + y + z = a$. We introduce new variables by putting:

$$x'_1 = x_1 + x_2 + \ldots + x_n; \qquad x'_2 = \frac{a(x_2 + \ldots + x_n)}{x_1 + x_2 + \ldots + x_n},$$

$$x'_3 = \frac{a(x_3 + \ldots + x_n)}{x_2 + \ldots + x_n};\ldots; \qquad x'_n = \frac{ax_n}{x_{n-1} + x_n},$$

whence it follows that

$$x_1 + \ldots + x_n = x'_1; \quad a(x_2 + \ldots + x_n) = x'_1 x'_2;$$

$$a^2(x_3 + \ldots + x_n) = x'_1 x'_2 x'_3;\ldots; \quad a^{n-1} x_n = x'_1 x'_2 \ldots x'_n.$$

Conversely, the old variables are given in terms of the new by the expressions

$$x_1 = \frac{x'_1(a - x'_2)}{a}; \qquad x_2 = \frac{x'_1 x'_2 (a - x'_3)}{a^2};\ldots;$$

$$x_{n-1} = \frac{x'_1 x'_2 \ldots x'_{n-1}(a - x'_n)}{a^{n-1}}; \qquad x_n = \frac{x'_1 x'_2 \ldots x'_n}{a^{n-1}}.$$

It follows at once from these formulae that tetrahedron (24) can be replaced by the n-dimensional cube:

$$0 < x'_1 < a;\ 0 < x'_2 < a;\ldots;\ 0 < x'_n < a. \tag{25}$$

2. We find the measure (volume) of the n-dimensional sphere with centre at the origin and radius r, defined by

$$x_1^2 + x_2^2 + \ldots + x_n^2 < r^2. \tag{26}$$

If a transformation of similitude is carried out with coefficient k, the volume of a cube is multiplied by k^n, and the radius of a sphere by k. It follows at once from this that the required measure v_n is a function of r only, of the form

$$v_n = C_n r^n, \tag{27}$$

where C_n is a numerical constant, different for different n. If sphere (26) is cut by a constant x_1 plane, it is clear from (26) that an $(n-1)$-dimensional sphere is obtained, the square of the radius of which is equal to $(r^2-x_1^2)$. The volume of the new sphere will be $C_{n-1}(r^2-x_1^2)^{(n-1)/2}$ by (27). The part of the n-dimensional sphere lying between the planes x_1 and x_1+dx_1 will have a volume $C_{n-1}(r^2-x_1^2)^{(n-1)/2}\,dx_1$, whence we have the following expression for v_n:

$$v_n = C_n r^n = C_{n-1}\int_{-r}^{+r}(r^2-x_1^2)^{\frac{n-1}{2}}\,dx_1,$$

or, on substituting $x_1 = r\cos\varphi$, we have the following relationship between C_n and C_{n-1}:

$$C_n = C_{n-1}\int_0^{\pi}\sin^n\varphi\,d\varphi = 2C_{n-1}\int_0^{\frac{\pi}{2}}\sin^n\varphi\,d\varphi, \tag{28}$$

where, as we know from [I, 100],

$$\int_0^{\frac{\pi}{2}}\sin^n\varphi\,d\varphi = \frac{(n-1)(n-3)\ldots 1}{n(n-2)\ldots 2}\,\frac{\pi}{2}\quad\text{for even } n,$$

$$\int_0^{\frac{\pi}{2}}\sin^n\varphi\,d\varphi = \frac{(n-1)(n-3)\ldots 2}{n(n-2)\ldots 3}\quad\text{for odd } n,$$

On replacing n by $(n-1)$ in (28), we get:

$$C_{n-1} = 2C_{n-2}\int_0^{\frac{\pi}{2}}\sin^{n-1}\varphi\,d\varphi.$$

It follows from the equations written that, for any integral n:

$$C_n = C_{n-2}\frac{2\pi}{n}. \tag{29}$$

But we know that $C_2=\pi$, $C_3=4\pi/3$. Hence we have from (29):

$$C_n = \frac{(2\pi)^{\frac{1}{2}n}}{n(n-2)\ldots 2}\quad\text{for even } n,$$

$$C_n = \frac{2^{\frac{1}{2}(n+1)}\,\pi^{\frac{1}{2}(n-1)}}{n(n-2)\ldots 1}\quad\text{for odd } n.$$

CHAPTER IV

VECTOR ANALYSIS AND FIELD THEORY

§ 10. Basic vector algebra

100. Addition and subtraction of vectors. The present chapter is mainly concerned with vector analysis. Since a number of specialized treatises are now available on the subject, we shall confine ourselves to the broad outlines, including only details directly connected with the previous matter and essential for our treatment of the foundations of mathematical physics.

We encounter two types of magnitude, scalars and vectors, when studying physical phenomena.

A scalar is a magnitude which, for a given choice of unit, is fully characterized by the number measuring it.

For instance, the temperature at any point of a heated body in space is characterized by a definite number, and we can, therefore, say that temperature is a scalar. Density, energy and potential are further examples of scalars.

Velocity may be taken as an example of a vector. To characterize a velocity, we need to know its direction as well as the number measuring its magnitude. In other words, we need to construct a vector: a straight line whose length is equal to the magnitude of the velocity on a given scale and the direction of which coincides with the direction of the velocity. *A vector is fully defined by its length and direction.* Force, acceleration and impulse are also vectors.

We return to our example of a heated body. The temperature u at any point of the body is characterized by a definite number and can be said to be a function of the points in space occupied by the body. If space is referred to a system XYZ of Cartesian coordinates, we can say that the scalar u is a function of the independent variables (x, y, z), defined in the domain of space occupied by the body. Here we have an example of a scalar field.

If a vector is defined at every point of a certain domain, we get a vector field, a suitable example being the electromagnetic field, where the electric and magnetic forces are defined at every point.

It becomes important in certain cases to know the point of application of a vector, i.e. the point in space at which the origin of the vector is situated. We shall be concerned in future, not so much with such tied vectors, as with free vectors, for which the point of application is arbitrary. We shall therefore look on two vectors as equal when they have equal magnitude (length) and the same direction.

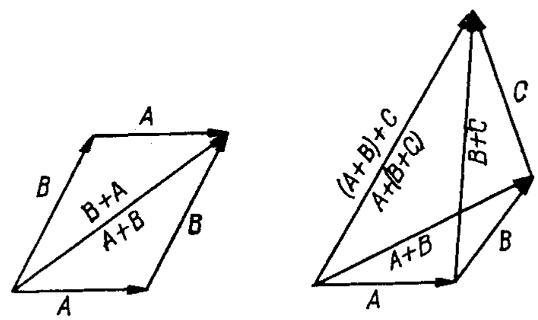

FIG. 84

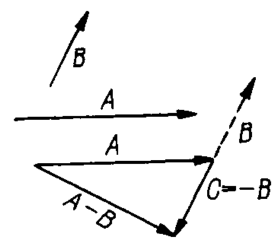

FIG. 85

Vectors will be written in heavy type $\mathbf{A}, \mathbf{B}, \ldots$, their respective magnitudes (lengths) being denoted as $|\mathbf{A}|, |\mathbf{B}|, \ldots$; letters in normal type are used for scalars.

Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ be three vectors. *We draw* $\mathbf{A}$ *with its origin at a given point O, we draw* $\mathbf{B}$ *from the terminus of* $\mathbf{A}$, *and* $\mathbf{C}$ *from the terminus of* $\mathbf{B}$. *The vector* $\mathbf{S}$, *with origin at the origin of the first vector and terminus at the terminus of the last vector, is called the sum of the given vectors:*

$$\mathbf{S} = \mathbf{A} + \mathbf{B} + \mathbf{C}.$$

Vector sums have the commutative and associative properties of ordinary sums, expressed by the formulae (Fig. 84):

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}; \quad \mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}.$$

If from the terminus of $\mathbf{A}$, *a vector* $\mathbf{C}$ *is drawn, equal in magnitude and opposite in direction to* $\mathbf{B}$, *the vector* $\mathbf{M}$ *with origin at the origin of* $\mathbf{A}$ *and terminus at the terminus of* $\mathbf{C}$ *is called the difference between* $\mathbf{A}$ *and* $\mathbf{B}$ (Fig. 85):

$$\mathbf{M} = \mathbf{A} - \mathbf{B}.$$

This vector is easily seen to be fully defined by the relationship:

$$\mathbf{B} + \mathbf{M} = \mathbf{A}.$$

In general we write $(-\mathbf{N})$ for the vector with the same magnitude but opposite direction to $\mathbf{N}$. The difference between $\mathbf{A}$ and $\mathbf{B}$ can then be defined as the sum of $\mathbf{A}$ and $(-\mathbf{B})$, i.e.

$$\mathbf{A} + (-\mathbf{B}) = \mathbf{A} - \mathbf{B}.$$

It is easily shown that the vector sum and difference thus defined are subject to the same rules as ordinary algebraic sums and differences, and we shall omit the proof.

The rule for vector addition has numerous applications in mechanics and physics. For instance, if a particle is subject to several types of motion, its final velocity is found by vector addition of the velocities due to each individual motion. Similarly, vector addition gives the resultant of several forces acting on a particle.

If the terminus of the final vector in a sum coincides with the origin of the initial vector, i.e. if the step line given by the above construction is closed, the sum is said to be zero:

$$\mathbf{A} + \mathbf{B} + \mathbf{C} = 0.$$

In particular, we obviously have:

$$\mathbf{A} + (-\mathbf{A}) = 0.$$

In general, a vector is said to be zero if its magnitude is zero; there is then no occasion for speaking of its direction.

101. Multiplication of a vector by a scalar. *Coplanar vectors. Given the vector* $\mathbf{A}$ *and the real number* a, *the product* $a\mathbf{A}$ *or* $\mathbf{A}a$ *is defined as the vector of magnitude* $|a| \cdot |\mathbf{A}|$, *its direction being the same as* $\mathbf{A}$ *if* $a > 0$, *and opposite to* $\mathbf{A}$ *if* $a < 0$. *If* $a = 0$, $a\mathbf{A}$ *also equals zero.*

Hence, if $\mathbf{A}$ and $\mathbf{B}$ are two vectors with the same or opposite directions, the relationship exists between them:

$$\mathbf{B} = n\mathbf{A},$$

which may be written more symmetrically as:

$$a\mathbf{A} + b\mathbf{B} = 0,$$

on setting $n = -a/b$.

Conversely, the existence of this relationship shows that $\mathbf{A}$ and $\mathbf{B}$ have the same or opposite directions.

Now let $\mathbf{A}$ and $\mathbf{B}$ be any two vectors whose directions are neither the same nor opposite. Let us draw two straight lines, parallel to the

given vectors, through an arbitrary point O (Fig. 86). These lines define a plane, which is parallel, not only to **A** and **B**, but to all $m\mathbf{A}$ and $n\mathbf{B}$, where m and n are any numbers, and to their sum, by the addition rule,

$$\mathbf{C} = m\mathbf{A} + n\mathbf{B}.$$

Conversely, any vector **C** parallel to the plane can be written in the form $m\mathbf{A} + n\mathbf{B}$. This may easily be seen by marking off **C** from O and representing it as the diagonal of a parallelogram with sides parallel to **A** and **B**. The above relationship may be written in the more symmetrical form:

$$a\mathbf{A} + b\mathbf{B} + c\mathbf{C} = 0,$$

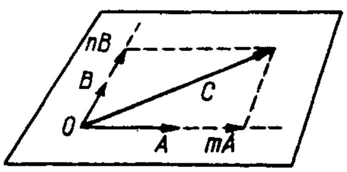

FIG. 86

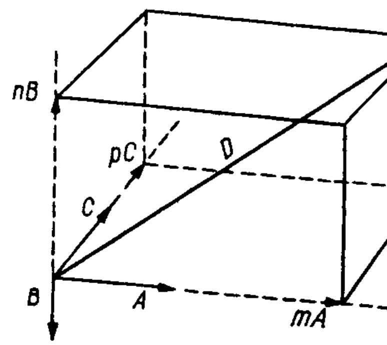

FIG. 87

and expresses the condition for three vectors to be coplanar, i.e. the fact that all three are parallel to the same plane. If **A** and **B** have the same or opposite directions, they are coplanar with any vector **C**, and we have to take $c = 0$ in the above relationship.

102. Resolution of a vector into three non-coplanar components. Let **A, B, C** be three vectors not in the same plane. Any vector can be represented as the diagonal of a parallelepiped, the three edges of which are parallel to **A**, **B**, **C**. Thus, any vector can be expressed in terms of three non-coplanar vectors (Fig. 87):

$$\mathbf{D} = m\mathbf{A} + n\mathbf{B} + p\mathbf{C}.$$

It follows from this that a relationship exists between any four vectors of the form:

$$a\mathbf{A} + b\mathbf{B} + c\mathbf{C} + d\mathbf{D} = 0.$$

If the first three vectors are in the same plane, we have to take $d = 0$.

The most important particular case of the above rule for resolving a vector into three others is when space is referred to rectangular coordinates XYZ and **A**, **B**, **C** are of unit length and are directed along

OX, OY, OZ respectively. They are then called *unit* vectors and are denoted by $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$.

Any vector $\mathbf{A}$ can be written as

$$\mathbf{A} = m\mathbf{i} + n\mathbf{j} + p\mathbf{k}. \tag{1}$$

If $\mathbf{A}$ is taken from the origin of coordinates, the numbers m, n, p are the coordinates of its terminus and give the projections of $\mathbf{A}$ on the axes. We shall in future refer to these projections as the components of $\mathbf{A}$ along the axes and write them as A_x, A_y, A_z. The above relationship can now be written as:

$$\mathbf{A} = A_x \mathbf{i} + A_y \mathbf{j} + A_z \mathbf{k}. \tag{2}$$

The projection of $\mathbf{A}$ along any direction n in space will be

$$A_n = |\mathbf{A}| \cos(n, \mathbf{A})$$

or, if we use the familiar expression of analytic geometry for the cosine of the angle between two directions:

$$\begin{aligned} A_n = |\mathbf{A}| [\cos(n, X) \cos(\mathbf{A}, X) + \cos(n, Y) \cos(\mathbf{A}, Y) + \\ + \cos(n, Z) \cos(\mathbf{A}, Z)] = A_x \cos(n, X) + \\ + A_y \cos(n, Y) + A_z \cos(n, Z). \end{aligned}$$

When vectors are added, their components are obviously added (the projections of the closing side of the vector polygon are equal to the sums of the projections of the components).

103. Scalar product. *The scalar product of two vectors* $\mathbf{A}$ *and* $\mathbf{B}$ *is defined as the scalar equal to the product of the magnitudes of the vectors multiplied by the cosine of the angle between them.*

We write the scalar product as $\mathbf{A} \cdot \mathbf{B}$, *so that*

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}| |\mathbf{B}| \cos(\mathbf{A}, \mathbf{B}). \tag{3}$$

It follows at once from the definition that

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A},$$

i.e. scalar products obey the *commutative law.*

If $\mathbf{A}$ and $\mathbf{B}$ are perpendicular, clearly

$$\mathbf{A} \cdot \mathbf{B} = 0.$$

In particular, we have for the unit vectors:

$$\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = 0.$$

If **A** and **B** have the same direction,

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}|\,|\mathbf{B}|,$$

whilst if their directions are opposite,

$$\mathbf{A}\cdot\mathbf{B} = -\,|\mathbf{A}|\,|\mathbf{B}|.$$

In particular,

$$\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2 = A_x^2 + A_y^2 + A_z^2 \tag{4}$$

and

$$\mathbf{i}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1. \tag{5}$$

The scalar product is given in terms of the components as follows:

$$\left.\begin{aligned}\mathbf{A}\cdot\mathbf{B} &= |\mathbf{A}|\,|\mathbf{B}|\cos(\mathbf{A},\mathbf{B}) = |\mathbf{A}|\,|\mathbf{B}|\,[\cos(\mathbf{A},X)\cos(\mathbf{B},X) + \\ &\quad + \cos(\mathbf{A},Y)\cos(\mathbf{B},Y) + \cos(\mathbf{A},Z)\cos(\mathbf{B},Z)] = \\ &= |\mathbf{A}|\cos(\mathbf{A},X)\,|\mathbf{B}|\cos(\mathbf{B},X) + |\mathbf{A}|\cos(\mathbf{B},Y)\,|\mathbf{B}|\cos(\mathbf{B},Y) + \\ &\quad + |\mathbf{A}|\cos(\mathbf{A},Z)\,|\mathbf{B}|\cos(\mathbf{B},Z) = A_xB_x + A_yB_y + A_zB_z,\end{aligned}\right\} \tag{6}$$

i.e. *the scalar product is equal to the sum of the products of corresponding components of the vectors.*

It may be noted that the obvious independence of the left-hand side of the above equation on the choice of axes implies the not-so-obvious independence of the right-hand side on the axes.

We have made use in (6) of the familiar expression of analytic geometry for the angle between two directions [102].

A scalar product may easily be seen to obey the distributive law, i.e. we have

$$(\mathbf{A}+\mathbf{B})\cdot\mathbf{C} = \mathbf{A}\cdot\mathbf{B} + \mathbf{B}\cdot\mathbf{C}. \tag{7}$$

We can use the expression just obtained for a scalar product to write:

$$\begin{aligned}(\mathbf{A}+\mathbf{B})\cdot\mathbf{C} &= (A_x+B_x)C_x + (A_y+B_y)C_y + (A_z+B_z)C_z = \\ &= (A_xC_x + A_yC_y + A_zC_z) + (B_xC_x + B_yC_y + B_zC_z) = \\ &= \mathbf{A}\cdot\mathbf{C} + \mathbf{B}\cdot\mathbf{C}.\end{aligned}$$

The distributive property gives us at once the more general formula

$$(\mathbf{A}_1+\mathbf{B}_1)\cdot(\mathbf{A}_2+\mathbf{B}_2) = \mathbf{A}_1\cdot\mathbf{A}_2 + \mathbf{A}_1\cdot\mathbf{B}_2 + \mathbf{B}_1\cdot\mathbf{A}_2 + \mathbf{B}_1\cdot\mathbf{B}_2. \tag{8}$$

which expresses the ordinary rule for removing brackets when cross-multiplying.

104. Vector products. We draw two vectors **A** and **B** through any point O of space and construct a parallelogram on them. The perpendicular through O to the plane of the parallelogram can have either

of two opposite directions. We take the direction with the property that, for an observer standing along it, the rotation of less than π required to turn the direction of $\mathbf{A}$ into that of $\mathbf{B}$ has the same sense as the rotation of $\pi/2$ needed to turn the positive direction of OX into that of OY for an observer standing along OZ. The direction of the perpendicular is shown in Fig. 88 for right- and left-handed systems of axes.

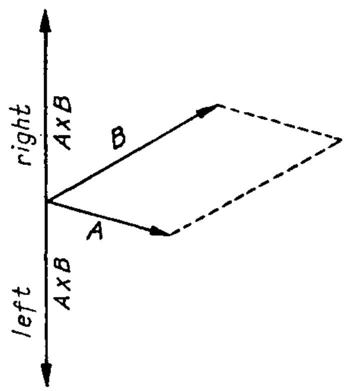

FIG. 88

The vector product of $\mathbf{A}$ *and* $\mathbf{B}$ *is defined as the vector of magnitude equal to the area of the parallelogram constructed from* $\mathbf{A}$ *and* $\mathbf{B}$ *and directed along the perpendicular as above defined to the plane of the parallelogram.*

The vector product is generally written symbolically $\mathbf{A} \times \mathbf{B}$. Its magnitude is, by definition:

$$|\mathbf{A}|\,|\mathbf{B}| \sin (\mathbf{A}, \mathbf{B}). \tag{9}$$

Its direction depends on the orientation of the axes, and if this is changed, its direction is reversed.

The vector product is zero if $\mathbf{A}$ and $\mathbf{B}$ have the same or opposite directions. In particular, obviously:

$$\mathbf{A} \times \mathbf{A} = 0.$$

We now consider the vector product of $\mathbf{B}$ and $\mathbf{A}$. Its magnitude is clearly the same as for the product of $\mathbf{A}$ and $\mathbf{B}$, whereas it has the opposite direction, since transposition of $\mathbf{A}$ and $\mathbf{B}$ means rotation of $\mathbf{B}$ only, and not $\mathbf{A}$, in the reverse sense. Thus:

$$\mathbf{B} \times \mathbf{A} = -\mathbf{A} \times \mathbf{B}, \tag{10}$$

whence it is clear that the commutative law does not apply; in fact, *interchange of the factors in a vector product changes its sign.*

We have the obvious relationships for the unit vectors:

$$\mathbf{i}\times\mathbf{i} = \mathbf{j}\times\mathbf{j} = \mathbf{k}\times\mathbf{k} = 0; \quad \mathbf{j}\times\mathbf{k} = \mathbf{i}; \quad \mathbf{k}\times\mathbf{i} = \mathbf{j}; \quad \mathbf{i}\times\mathbf{j} = \mathbf{k}. \tag{11}$$

We next express the components of the vector product $\mathbf{P} = \mathbf{A} \times \mathbf{B}$ in terms of the components of $\mathbf{A}$ and $\mathbf{B}$. We can write, since $\mathbf{A} \times \mathbf{B}$ is perpendicular to $\mathbf{A}$ and $\mathbf{B}$:

$$P_x A_x + P_y A_y + P_z A_z = 0, \quad P_x B_x + P_y B_y + P_z B_z = 0.$$

We now use the following elementary algebraic lemma.

LEMMA. *The solution of the two homogeneous equations in three variables*

$$ax + by + cz = 0; \quad a_1 x + b_1 y + c_1 z = 0$$

has the form

$$x = \lambda(bc_1 - cb_1); \quad y = \lambda(ca_1 - ac_1); \quad z = \lambda(ab_1 - ba_1),$$

where λ is an arbitrary factor and it is assumed that at least one of the differences in brackets is not zero.

We leave the proof of this simple lemma to the reader; it gives us †:

$$P_x = \lambda(A_y B_z - A_z B_y); \quad P_y = \lambda(A_z B_x - A_x B_z);$$
$$P_z = \lambda(A_x B_y - A_y B_x),$$

where λ is a coefficient of proportionality which must still be defined.

For this, we use an important additional identity, generally known as *Lagrange's identity:*

$$(a^2 + b^2 + c^2)(a_1^2 + b_1^2 + c_1^2) - (aa_1 + bb_1 + cc_1)^2 =$$
$$= (bc_1 - cb_1)^2 + (ca_1 - ac_1)^2 + (ab_1 - ba_1)^2, \tag{12}$$

which is easily verified by removing the brackets on both sides. We also notice that $(P_x^2 + P_y^2 + P_z^2)$ is the square of the length of P, i.e.

$$\lambda^2 [(A_y B_z - A_z B_y)^2 + (A_z B_x - A_x B_z)^2 + (A_x B_y - A_y B_x)^2] =$$
$$= |\mathbf{A}|^2 |\mathbf{B}|^2 \sin^2 (\mathbf{A}, \mathbf{B}).$$

We apply Lagrange's identity to the left-hand side of this last equation, which gives us

$$\lambda^2 ([A_x^2 + A_y^2 + A_z^2)(B_x^2 + B_y^2 + B_z^2) - (A_x B_x + A_y B_y + A_z B_z)^2] =$$
$$= |\mathbf{A}|^2 |\mathbf{B}|^2 \sin^2 (\mathbf{A}, \mathbf{B}),$$

or, on using (4) and (6):

$$\lambda^2 [\,|\mathbf{A}|^2 |\mathbf{B}|^2 - |\mathbf{A}|^2 |\mathbf{B}|^2 \cos^2 (\mathbf{A}, \mathbf{B})] = |\mathbf{A}|^2 |\mathbf{B}|^2 \sin^2 (\mathbf{A}, \mathbf{B}),$$

whence it follows at once that $\lambda = \pm 1$.

We finally show that $\lambda = +1$. We subject $\mathbf{A}$ and $\mathbf{B}$ to continuous deformation, such that $\mathbf{A}$ coincides with the unit vector $\mathbf{i}$, and $\mathbf{B}$ with $\mathbf{j}$. The deformation can be carried out in such a way that

† If all three of the differences in brackets are zero, the angle between $\mathbf{A}$ and $\mathbf{B}$ is 0 or π, and $\mathbf{A} \times \mathbf{B} = 0$, i.e. $P_x = P_y = P_z = 0$.

A and **B** do not vanish and do not become parallel to each other. Then $\mathbf{A} \times \mathbf{B}$ will not vanish and will finally become

$$\mathbf{i} \times \mathbf{j} = \mathbf{k},$$

since **A** coincides with **i**, and **B** with **j**.

Since the change is continuous and λ can only have two values (± 1), we can say that λ will not change during the deformation, and hence it must keep its final value. But we have after deformation:

$$A_x = 1; \quad A_y = A_z = 0; \quad B_y = 1; \quad B_x = B_z = 0;$$
$$P_z = 1; \quad P_x = P_y = 0,$$

and we can conclude from the relationship

$$P_z = \lambda(A_x B_y - A_y B_x)$$

that $\lambda = +1$.

We thus get the following expressions for the components of $\mathbf{A} \times \mathbf{B}$:

$$A_y B_z - A_z B_y; \quad A_z B_x - A_x B_z; \quad A_x B_y - A_y B_x. \tag{13}$$

By means of these expressions, the reader may easily verify that vector products obey the distributive law, i.e. that:

$$(\mathbf{A} + \mathbf{B}) \times \mathbf{C} = \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C}. \tag{14}$$

Hence we find, on using (10):

$$\mathbf{C} \times (\mathbf{A} + \mathbf{B}) = \mathbf{C} \times \mathbf{A} + \mathbf{C} \times \mathbf{B},$$

which leads on to the more general formula:

$$(\mathbf{A}_1 + \mathbf{A}_2) \times (\mathbf{B}_1 + \mathbf{B}_2) =$$
$$= \mathbf{A}_1 \times \mathbf{B}_1 + \mathbf{A}_1 \times \mathbf{B}_2 + \mathbf{A}_2 \times \mathbf{B}_1 + \mathbf{A}_2 \times \mathbf{B}_2, \tag{15}$$

which is the exact analogue of (8) for scalar products.

105. The relationship between scalar and vector products. We form the scalar product of the vector **A** with the vector product $\mathbf{N} = \mathbf{B} \times \mathbf{C}$:

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}).$$

The magnitude of **N** is given by the area of the parallelogram formed from **B** and **C**. But

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{A} \cdot \mathbf{N} = |\mathbf{A}|\,|\mathbf{N}| \cos(\mathbf{A}, \mathbf{N}),$$

and the scalar product can therefore be looked on as the product of area $|\mathbf{N}|$ of the parallelogram and the projection of **A** on the direction

of $\mathbf{N}$, which is perpendicular to the area, i.e. *the scalar product* $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ *expresses the volume of the parallelepiped formed from* $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. Its sign depends on the orientation of the coordinate system. It is easily seen to be $(+)$ if the set of vectors $\mathbf{B}$, $\mathbf{C}$, $\mathbf{A}$ or, what amounts to the same thing, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ has the same orientation as the axes. We can prove this by the method of continuous deformation used above†.

When referring to the volume of the parallelepiped, we took as its base the parallelogram formed from $\mathbf{B}$ and $\mathbf{C}$. We might equally well have taken the base as the parallelogram formed from $\mathbf{C}$ and $\mathbf{A}$ or from $\mathbf{A}$ and $\mathbf{B}$. We thus have the following relationships:

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}). \tag{16}$$

It only remains to notice the signs of these scalar products; they are the same, since the sets $(\mathbf{A}, \mathbf{B}, \mathbf{C})$, $(\mathbf{B}, \mathbf{C}, \mathbf{A})$, and $(\mathbf{C}, \mathbf{A}, \mathbf{B})$ have the same orientation. The two latter sets are acquired from the first by a cyclic change. A different order changes the sign, e.g.

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = -\,\mathbf{B}\cdot(\mathbf{A}\times\mathbf{C}). \tag{17}$$

If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are coplanar, the volume of the parallelepiped is zero, i.e.

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = 0. \tag{18}$$

This equation is the necessary and sufficient condition for three vectors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *to be coplanar.*

We now consider the vector product of $\mathbf{A}$ and $\mathbf{B}\times\mathbf{C}$, i.e

$$\mathbf{D} = \mathbf{A}\times(\mathbf{B}\times\mathbf{C}).$$

Since $\mathbf{D}$ is perpendicular to $\mathbf{B}\times\mathbf{C}$, it is coplanar with $\mathbf{B}$ and $\mathbf{C}$, and hence [101]:

$$\mathbf{D} = m\mathbf{B} + n\mathbf{C}; \tag{19}$$

but $\mathbf{D}$ is also perpendicular to $\mathbf{A}$, and hence [103]:

$$\mathbf{A}\cdot\mathbf{D} = m\mathbf{A}\cdot\mathbf{B} + n\mathbf{A}\cdot\mathbf{C} = 0,$$

so that

$$m = \mu\mathbf{A}\cdot\mathbf{C};\quad n = -\,\mu\mathbf{A}\cdot\mathbf{B},$$

† The dependence of the sign of $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ on the orientation of the axes comes from the dependence of $\mathbf{B}\times\mathbf{C}$ on the orientation. Hence $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is not an ordinary scalar, the magnitude of which has to be independent of the choice of axes. In general, quantities that depend on the axes only to the extent that a change of orientation implies a change of sign are called ***pseudoscalars***.

and finally

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{D} = \mu\{(\mathbf{A}\cdot\mathbf{C})\,\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\,\mathbf{C}\},$$

where it only remains to find the coefficient of proportionality μ. For this, it is enough to compare the components along one of the axes of the vectors on the left and right-hand sides of the last equation. We take OX parallel to $\mathbf{A}$ and find the components along OZ. On noting that with this choice of axes

$$A_x = |\mathbf{A}| = a; \quad A_y = A_z = 0,$$

we have for the left-hand side [104]

$$D_z = A_x\,(\mathbf{B}\times\mathbf{C})_y = a(B_z\,C_x - B_x\,C_z),$$

and for the right [103]

$$\mu(aC_x\,B_z - aB_x\,C_z),$$

on equating which, we get $\mu = 1$.

This leads us to the following expression:

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{C}\cdot\mathbf{A})\,\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\,\mathbf{C}. \tag{20}$$

As a corollary of this, we can *resolve a vector* $\mathbf{B}$ *into two component vectors, parallel and perpendicular to a given vector* $\mathbf{A}$. We put $\mathbf{C} = \mathbf{A}$ in (20) and re-write it as

$$(\mathbf{A}\cdot\mathbf{A})\,\mathbf{B} = (\mathbf{A}\cdot\mathbf{B})\,\mathbf{A} - \mathbf{A}\times(\mathbf{A}\times\mathbf{B})$$

or

$$\mathbf{B} = \mathbf{B}' + \mathbf{B}'', \tag{21}$$

where

$$\mathbf{B}' = \frac{\mathbf{A}\cdot\mathbf{B}}{\mathbf{A}\cdot\mathbf{A}}\,\mathbf{A}; \quad \mathbf{B}'' = -\,\frac{\mathbf{A}\times(\mathbf{A}\times\mathbf{B})}{\mathbf{A}\cdot\mathbf{A}},$$

which gives the required result, since clearly, $\mathbf{B}'$ is parallel to, and $\mathbf{B}''$ perpendicular to, $\mathbf{A}$.

106. The velocities at points of a rotating rigid body; the moment of a vector. Vector products have important applications in mechanics, primarily in the *dynamics of a rigid body*.†

We first consider a rigid body rotating about a fixed axis (L). In this case, every point M of the body has a velocity $\mathbf{v}$ which is equal in magnitude to the product of the distance PM of M from the axis of rotation (Fig. 89) and the angular velocity ω of the rotation, and which has a direction perpendicular to the plane containing M and the axis of rotation. The velocity $\mathbf{v}$ can be represented geometrically as follows. We choose the direction along (L) with respect

† We use a right-handed system of axes in future.

to which the rotation is counter-clockwise and call this positive. If we mark off a length equal to ω in this direction from an arbitrary point A of the axis, we get a vector $\mathbf{o}$ called the angular velocity vector. We let $\mathbf{r}$ denote the vector along $\overline{AM}$ and use the definition of vector product; this gives us the simple expression for $\mathbf{v}$:

$$\mathbf{v} = \mathbf{o} \times \mathbf{r},$$

since the magnitude of $\mathbf{o} \times \mathbf{r}$ is equal to

$$|\mathbf{r}|\,|\mathbf{o}| \sin(\mathbf{r}, \mathbf{o}) = \omega \cdot |\overrightarrow{MA}| \cdot \sin\varphi = \omega \cdot |MP| = |\mathbf{v}|,$$

whilst its direction is the same as that of $\mathbf{v}$.

We know from kinematics that, in the case of the motion of a rigid body about a fixed point O, the velocity of any point of the body at any given instant is the same as it would be if the body rotated about an axis (the instantaneous axis) through the point O with an angular velocity ω (the instantaneous angular velocity); generally speaking, the position of the axis and the value of ω change with time. By the above, *the velocity of a point of a rigid body is given at any given instant by the vector product of the instantaneous angular velocity and the vector* **OM**.

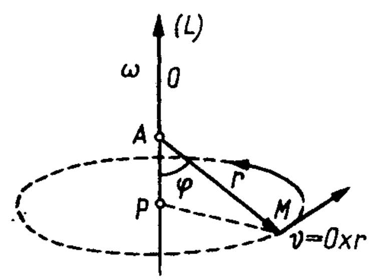

FIG. 89

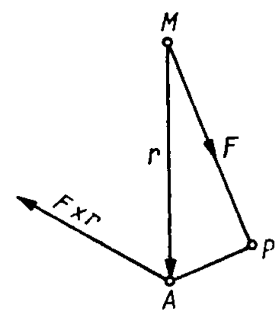

FIG. 90

We take a second example. Let a force represented by the vector $\mathbf{F}$ be applied at the point M, and let A be a given point of space (Fig. 90).

We define the moment of the force $\mathbf{F}$ *about the point* A *as the vector product* $\mathbf{F} \times \mathbf{r}$, *where* $\mathbf{r}$ *is the vector with origin at* M *and terminus at* A.

Let AP be the perpendicular from A to the line along which $\mathbf{F}$ lies. We get from the right-angled triangle AMP:

$$|\overrightarrow{AP}| = |\mathbf{r}|\,|\sin(\mathbf{r}, \mathbf{F})|$$

so that the *magnitude of the moment of the force* $\mathbf{F}$ about A is

$$|\mathbf{r}|\,|\mathbf{F}|\,|\sin(\mathbf{r}, \mathbf{F})| = |\mathbf{F}|\,|\overrightarrow{AP}|,$$

i.e. *it is equal to the product of the magnitude of the force and the distance of the point* A *from the line along which the force acts.* The direction of the moment is found by using the above rule for the direction of a vector product.

It follows from what has been said that the moment of a force does not vary with movement of the point of application M along the line of the force. The

definition of the moment of a force about a point can evidently be generalized for the case of any vector.

We now find the components of the moment. Let (a, b, c) be the coordinates of A and (x, y, z) the coordinates of M. The components of $\mathbf{r}$ are:

$$a - x, \quad b - y, \quad c - z.$$

We use the expressions for the components of a vector product [104] to obtain the components of the moment as follows:

$$(y - b) F_z - (z - c) F_y; \quad (z - c) F_x - (x - a) F_z; \quad (x - a) F_y - (y - b) F_x.$$

Returning to the example of a rigid body rotating about an axis, we can say that the velocity of a point M of the body is equal to the moment of the angular velocity vector about M. If we take the coordinates of M as (x, y, z), whilst (x_0, y_0, z_0) is the origin of the angular velocity vector, with components O_x, O_y, O_z, we have the following expressions for the components of the velocity of M:

$$(z - z_0) O_y - (y - y_0) O_z; \; (x - x_0) O_z - (z - z_0) O_x; \; (y - y_0) O_x - (x - x_0) O_y.$$

We now define the moment of a vector about an axis. Let Δ be a straight line in space to which a definite direction is assigned (the axis).

The moment of a vector $\mathbf{F}$ *about the axis* Δ *is defined as the algebraic value of the projection on the axis of the moment of* $\mathbf{F}$ *about any point* A *of* Δ.†

The definition is justified by showing the independence of the projection of the position of A on Δ. We take Δ as axis OZ and let $(0, 0, c)$ be the coordinates of A and (x, y, z) the coordinates of the origin M of $\mathbf{F}$. With this choice of axes, the projection on Δ of the moment of $\mathbf{F}$ about A is the same as its component along OZ, and is equal, by the above formula, to

$$xF_y - yF_x,$$

since $a = b = 0$. This is independent of c, i.e. of the position of A on Δ.

§ 11. Field theory

107. Differentiation of vectors. We extend differentiation to the case of a variable vector $\mathbf{A}(\tau)$, depending on a numerical parameter τ. We shall take the vector from a fixed point, say the coordinate origin O (Fig. 91). As τ varies, the terminus of the vector $\mathbf{A}(\tau)$ describes a certain curve (L). Let the positions of the vector for the values of the parameter $(\tau + \Delta\tau)$ and τ be respectively OM_1 and OM. The straight line MM_1 corresponds to the difference $\mathbf{A}(\tau + \Delta\tau) - \mathbf{A}(\tau)$, and the ratio

$$\frac{\mathbf{A}(\tau + \Delta\tau) - \mathbf{A}(\tau)}{\Delta\tau}$$

† The moment of a vector about a point is a vector, whilst the moment about an axis is not.

gives us a vector parallel to MM_1. The limiting position of this vector as $\Delta\tau \to 0$, if it exists, represents the derivative

$$\frac{\mathrm{d}\mathbf{A}(\tau)}{\mathrm{d}\tau} = \lim_{\Delta\tau\to 0} \frac{\mathbf{A}(\tau+\Delta\tau)-\mathbf{A}(\tau)}{\Delta\tau}. \tag{22}$$

This derivative is evidently a vector directed along the tangent to the curve (L) at M. It also depends on τ, and its derivative with respect to τ gives a second derivative $\mathrm{d}^2\mathbf{A}(\tau)/\mathrm{d}\tau^2$ and so on.

We can write $\mathbf{A}(\tau)$ in terms of the unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$:

$$\mathbf{A}(\tau) = A_x(\tau)\mathbf{i} + A_y(\tau)\mathbf{j} + A_z(\tau)\mathbf{k}.$$

Definition (22) now gives

$$\frac{\mathrm{d}\mathbf{A}(\tau)}{\mathrm{d}\tau} = \frac{\mathrm{d}A_x(\tau)}{\mathrm{d}\tau}\mathbf{i} + \frac{\mathrm{d}A_y(\tau)}{\mathrm{d}\tau}\mathbf{j} + \frac{\mathrm{d}A_z(\tau)}{\mathrm{d}\tau}\mathbf{k} \tag{23}$$

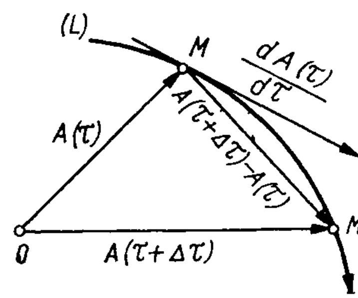

FIG. 91

and in general

$$\frac{\mathrm{d}^m\mathbf{A}(\tau)}{\mathrm{d}\tau^m} = \frac{\mathrm{d}^m A_x(\tau)}{\mathrm{d}\tau^m}\mathbf{i} + \frac{\mathrm{d}^m A_y(\tau)}{\mathrm{d}\tau^m}\mathbf{j} + \frac{\mathrm{d}^m A_z(\tau)}{\mathrm{d}\tau^m}\mathbf{k}, \tag{23_1}$$

i.e. *differentiation of a vector amounts to differentiation of its components.*

The familiar rule for differentiation of a product may be extended to the product of a scalar and a vector and to scalar and vector products, in accordance with the formulae:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\{f(\tau)\mathbf{A}(\tau)\} = \frac{\mathrm{d}f(\tau)}{\mathrm{d}\tau}\mathbf{A}(\tau) + f(\tau)\frac{\mathrm{d}\mathbf{A}(\tau)}{\mathrm{d}\tau} \tag{24}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{A}(\tau)\cdot\mathbf{B}(\tau) = \frac{\mathrm{d}\mathbf{A}(\tau)}{\mathrm{d}\tau}\cdot\mathbf{B}(\tau) + \mathbf{A}(\tau)\cdot\frac{\mathrm{d}\mathbf{B}(\tau)}{\mathrm{d}\tau} \tag{24_1}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{A}(\tau)\times\mathbf{B}(\tau) = \frac{\mathrm{d}\mathbf{A}(\tau)}{\mathrm{d}\tau}\times\mathbf{B}(\tau) + \mathbf{A}(\tau)\times\frac{\mathrm{d}\mathbf{B}(\tau)}{\mathrm{d}\tau}, \tag{24_2}$$

where $f(\tau)$ is a scalar, and $\mathbf{A}(\tau)$, $\mathbf{B}(\tau)$ are vectors, depending on τ. We verify say (24_1); the left-hand side may be written:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\{A_x(\tau)B_x(\tau) + A_y(\tau)B_y(\tau) + A_z(\tau)B_z(\tau)\} =$$

$$= \frac{\mathrm{d}A_x(\tau)}{\mathrm{d}\tau}B_x(\tau) + \frac{\mathrm{d}A_y(\tau)}{\mathrm{d}\tau}B_y(\tau) + \frac{\mathrm{d}A_z(\tau)}{\mathrm{d}\tau}B_z(\tau) + A_x(\tau)\frac{\mathrm{d}B_x(\tau)}{\mathrm{d}\tau} +$$

$$+ A_y(\tau)\frac{\mathrm{d}B_y(\tau)}{\mathrm{d}\tau} + A_z(\tau)\frac{\mathrm{d}B_z(\tau)}{\mathrm{d}\tau}.$$

The same result may be easily seen to be obtained for the right-hand side. Of course it is assumed that the derivatives exist. The existence

of the derivatives in the factors of (24), (24_1), (24_2) implies the existence of the derivative of each product [cf. I, 47]. The proof of the ordinary rule for differentiating a vector sum is purely elementary. If a point M moves along a curve (L), the radius vector $\mathbf{r}$ of M is a function of time t. We get the velocity vector of the point on differentiating the radius vector with respect to t:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{ds}{dt} \cdot \frac{d\mathbf{r}}{ds} . \tag{25}$$

The length of this vector is given by the derivative of the path s with respect to time t, whilst its direction is tangential to (L). The velocity vector also depends on time and we can differentiate it to get the acceleration vector $\mathbf{w} = d\mathbf{v}/dt$.

If we take the length s of the curve as the independent variable, *the derivative of* $\mathbf{r}$ *with respect to* s *is represented by the unit vector along the tangent* $\mathbf{t}=d\mathbf{r}/ds$. We saw in [I, 70] that $\sqrt{\Delta x^2+\Delta y^2}/\Delta s \to 1$, that is, the ratio of length of chord to length of corresponding arc tends to unity. The same is evidently true for curves in space [I, 160]. It follows at once from this and definition (22) with $\tau = s$ that the length of the above tangential vector is in fact unity.

108. Scalar field and gradient. If a physical quantity has a definite value at every point of the whole or part of space, a field of the quantity is defined. The field is said to be scalar, if the quantity is scalar (temperature, pressure, electrostatic potential), or vector [100], if the quantity is a vector (velocity, force).

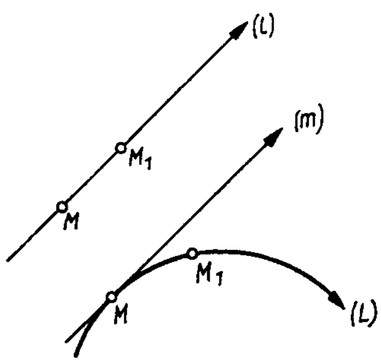

FIG. 92

We start with a scalar field, which is defined simply by defining a function of a point $U(M) = U(x, y, z)$.

A heated body, for instance, gives a scalar temperature field. The temperature $U(M)$ has a definite value at any point M of the body, and can vary from point to point.

We take a definite point and draw a straight line through it with an assigned direction (l) (Fig. 92). We consider the value of $U(M)$ at the point M and at a neighbouring point M_1 on (l). The limit of the ratio

$$\frac{U(M_1) - U(M)}{\overline{MM_1}}$$

is called the *derivative of the function $U(M)$ with respect to the direction* (l) and is written as:

$$\frac{\partial U(M)}{\partial l} = \lim_{M_1 \to M} \frac{U(M_1) - U(M)}{\overline{MM_1}}. \tag{26}$$

This derivative characterizes the rapidity of variation of $U(M)$ at the point M in the direction (l). The function thus has an unlimited number of derivatives at any given point; but it is easily shown that the derivative with respect to any given direction is expressible in terms of the derivatives with respect to the three mutually perpendicular directions X, Y, Z in accordance with:

$$\frac{\partial U(M)}{\partial l} = \frac{\partial U(M)}{\partial x} \cos(l, X) + \frac{\partial U(M)}{\partial y} \cos(l, Y) + \\ + \frac{\partial U(M)}{\partial z} \cos(l, Z). \tag{27}$$

We remark first of all that we could have formed derivative (26) by taking any directed curve (L) through the point M instead of a straight line (Fig. 92). Instead of (26), we should have had to consider the limit

$$\lim_{M_1 \to M} \frac{U(M_1) - U(M)}{\smile MM_1}.$$

This is, in fact, clearly the derivative of $U(M)$ with respect to the length of arc s of the curve (L) and we can write, on using the rule for differentiation of a function of a function:

$$\lim_{M_1 \to M} \frac{U(M_1) - U(M)}{\smile MM_1} = \frac{\partial U(M)}{\partial x} \cdot \frac{dx}{ds} + \frac{\partial U(M)}{\partial y} \cdot \frac{dy}{ds} + \\ + \frac{\partial U(M)}{\partial z} \cdot \frac{dz}{ds}. \tag{28}$$

But we know from [I, 160] that dx/ds, dy/ds, dz/ds are the direction-cosines of the tangent to (L) at M and when (L) is a straight line, we again get (27). Furthermore, (28) shows that the derivative with respect to the curve is the same as the derivative with respect to the tangent (m) to the curve at M.

We now bring in the idea of level surfaces of a scalar field. These surfaces are characterized by the condition that $U(M)$ has the same constant value C at all points of the surface. On assigning different numerical values to the constant, we get a family of level surfaces $U(M) = C$, with a definite surface passing through any given point

of space. In the case of a heated body, the level surfaces are those of constant temperature. Let (S) be the level surface passing through the point M (Fig. 93). We take three mutually perpendicular directions through M: the normal (n) to (S), and two directions (t_1), (t_2) lying in the tangent plane. The directions (t_1), (t_2) will be tangential to curves (L_1), (L_2) respectively, lying on the level surface. Since $U(M)$ is constant along these curves, we have:

$$\frac{\partial U(M)}{\partial t_1} = \frac{\partial U(M)}{\partial t_2} = 0, \tag{29}$$

We now take any direction (l). We use (27) with (n), (t_1) and (t_2) as the three mutually perpendicular directions and take (29) into account, which gives:

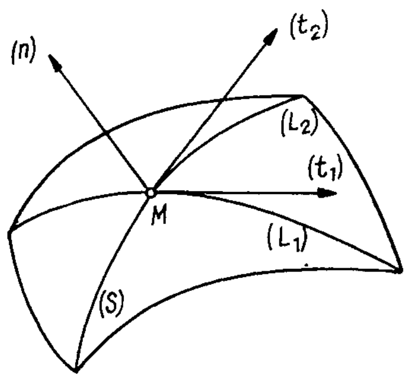

FIG. 93

$$\frac{\partial U(M)}{\partial l} = \frac{\partial U(M)}{\partial n} \cos(l, n). \tag{30}$$

If we draw a vector along (n) of algebraic magnitude $\partial U(M)/\partial n$, its projection on any direction (l) gives the derivative $\partial(U)M/\partial l$, by (30).

The vector thus drawn is called the *gradient of* $U(M)$, i.e. *the gradient of a scalar field is, by definition, the vector field obtained as follows: the vector at any point is along the normal to the corresponding level surface, whilst its algebraic magnitude is equal to the derivative of* $U(M)$ *with respect to the normal.* The gradient of the scalar field $U(M)$ is written grad $U(M)$, and we can write (30) as:

$$\frac{\partial U(M)}{\partial l} = \operatorname{grad}_l U(M), \tag{31}$$

where $\operatorname{grad}_l U(M)$ is the projection of the vector grad $U(M)$ on (l).

It is easily seen that the choice of the direction of normal (n) to the level surface (S) does not affect the direction of grad $U(M)$. The latter vector is always in the normal direction to (S) in which $U(M)$ is increasing.

Examples. **1.** The gravitational field discussed in [87] leads to a scalar field of the gravitational potential

$$U(M) = \iiint\limits_{(v)} \frac{\mu(M_1)\,dv}{r},$$

where $\mu(M_1)$ is the density of the matter occupying volume (v) and r is the distance of the point M from the variable point M_1 of integration. We had the following expressions for the components of the gravitational force:

$$F_x = \frac{\partial U(M)}{dx}, \quad F_y = \frac{\partial U(M)}{\partial y}, \quad F_z = \frac{\partial U(M)}{\partial z},$$

where F_x, F_y, F_z are the components of the vector $\mathbf{F}$. It follows at once that, in general, $F_l = \partial U(M)/\partial l$, i.e. the vector field of the gravitational force is the gradient of the potential $U(M)$. The work done by the force is given by

$$\int_{(A)}^{(B)} F_x\,dx + F_y\,dy + F_z\,dz = \int_{(A)}^{(B)} dU(M) = U(B) - U(A),$$

i.e. the work is given by the potential difference between points A and B.

Every conservative field of force, i.e. every field for which $\mathbf{F} = \operatorname{grad} U(M)$, clearly has this property. We often call $-U(M)$ the potential, instead of $U(M)$ itself.

2. If different points of a body have different temperatures $U(M)$, movement of heat will take place in the field from hotter to cooler points. Let dS be a small element about M of any surface through M. We find from thermodynamics that the quantity of heat crossing the element dS in time dt is proportional to $dt\,dS$ and the normal temperature derivative $\partial U(M)/\partial n$, i.e.

$$\Delta Q = k\,dt\,dS \left|\frac{\partial U(M)}{\partial n}\right|, \tag{32}$$

where k is the coefficient of proportionality called the internal thermal con ductivity and (n) is the normal to dS.

We now draw the vector $-k \operatorname{grad} U(M)$, called the *heat flow vector;* the $(-)$ sign is used because heat flows from points of higher to lower temperature, whilst grad $U(M)$ is along the normal to the level surface in the direction of increasing $U(M)$. We can say by (32) that the quantity of heat ΔQ passing in time dt through the element dS is given by:

$$\Delta Q = k\,dt\,dS\,|\operatorname{grad}_n U(M)|. \tag{33}$$

109. Vector fields. Curl and divergence. We now consider a vector field $\mathbf{A}(M)$, obtained when a vector $\mathbf{A}(M)$ is defined in magnitude and direction at every point of the whole or part of space occupied by the field. In fluid flow, for instance, we have a vector field of the velocity $\mathbf{v}$ at any given instant.

We define a vector line of the field as a curve (L) such that the tangent at every point has the direction of $\mathbf{A}(M)$ (Fig. 94). It may easily be seen, as in [22], that a vector line has a differential equation of the form

$$\frac{dx}{A_x} = \frac{dy}{A_y} = \frac{dz}{A_z}, \tag{34}$$

where the components A_x, A_y, A_z are known functions of x, y, z. If the conditions of the existence and uniqueness theorem are observed, it follows from this that one and only one vector line passes through any point M†. If we draw all the vector lines through all the points of a piece of surface (S), their aggregate gives us a vector tube (Fig. 94).

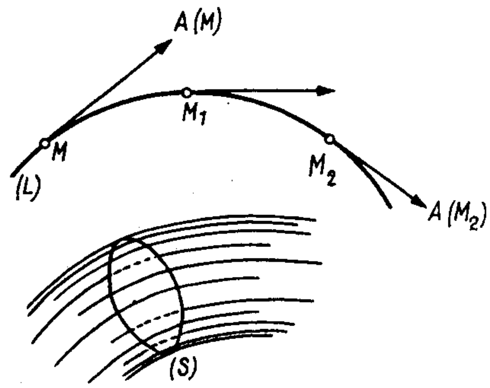

FIG. 94

Let (v) be a volume in the field bounded by the surface (S) and let (n) be the direction of the outward normal to (S) from (v). We find on applying Ostrogradskii's formula [63] to functions A_x, A_y, A_z:

$$\iiint\limits_{(v)} \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right) dv =$$
$$= \iint\limits_{(S)} [A_x \cos(n, X) + A_y \cos(n, Y) + A_z \cos(n, Z)]\, dS$$

or [102]

$$\iiint\limits_{(v)} \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right) dv = \iint\limits_{(S)} A_n\, dS. \tag{35}$$

The surface integral on the right is usually called the *flux of the field through the surface*. Its physical significance is explained later. The integrand of the volume integral is called the divergence of the vector field and is written

$$\operatorname{div} \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\,. \tag{36}$$

† The conditions of the theorem are certainly satisfied if A_x, A_y, A_z are continuous and have continuous derivatives, whilst $\mathbf{A}(M)$ differs from zero at M, i.e. at least one component differs from zero [51].

We can thus write Ostrogradskii's formula as

$$\iiint\limits_{(v)} \operatorname{div} \mathbf{A}\, \mathrm{d}v = \iint\limits_{(S)} A_n \,\mathrm{d}S, \tag{37}$$

i.e. *the volume integral of the divergence is equal to the flux of the field through the surface of the volume.* Definition (36) of divergence is related to the choice of axes X, Y, Z, but an independent definition is easily obtained with the aid of (37). Let (S_1) be the surface of a small volume (v_1) surrounding the point M. If we apply (37) and use the mean value theorem [61], we can write

$$\operatorname{div} \mathbf{A}\Big|_{M_1} \cdot v_1 = \iint\limits_{(S_1)} A_n \,\mathrm{d}S, \quad \text{i.e.} \quad \operatorname{div} \mathbf{A}\Big|_{M_1} = \frac{\iint\limits_{(S_1)} A_n \,\mathrm{d}S}{v_1},$$

where the value of div $\mathbf{A}$ is taken at a point M_1 of volume (v_1), the magnitude of which is denoted by v_1. On indefinite contraction of the volume to the point M, M_1 tends to M, and the above expression gives in the limit the value of the divergence at M:

$$\operatorname{div} \mathbf{A} = \lim_{(v_1)\to M} \frac{\iint\limits_{(S_1)} A_n \,\mathrm{d}S}{v_1}, \tag{38}$$

i.e. *the divergence of the field at the point M is the limit of the ratio of the flux through a small surface surrounding M to the volume enclosed by this surface.*

The above discussion shows that every vector field $\mathbf{A}$ gives rise to a scalar field div $\mathbf{A}$, i.e. the field of its divergence. We show next that, by using Stokes' theorem, we can generate a unique vector field from the initial field $\mathbf{A}$. We take

$$P = A_x; \quad Q = A_y; \quad R = A_z,$$

and write Stokes' formula [70] as:

$$\int\limits_{(l)} A_x \,\mathrm{d}x + A_y \,\mathrm{d}y + A_z \,\mathrm{d}z = \iint\limits_{(S)} \Big[\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\cos(n, X) +$$

$$+ \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\cos(n, Y) + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\cos(n, Z)\Big]\mathrm{d}S. \tag{39}$$

Let $\mathbf{ds}$ be a directed element of arc of the curve (l), i.e. an element of arc considered as a small vector. Its components on the axes will be $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$, and the expression under the line integral sign consists of the scalar product $\mathbf{A} \cdot \mathbf{ds}$, i.e. is equal to $A_s\,\mathrm{d}s$, where A_s is the projection $\mathbf{A}$ on the tangent to (l).

We introduce another vector, with components equal to the differences appearing under the double integral sign. This vector forms the new vector field and is called the curl of field $\mathbf{A}$, being written curl $\mathbf{A}$ thus:

$$\operatorname{curl}_x \mathbf{A} = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}; \quad \operatorname{curl}_y \mathbf{A} = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}; \quad \operatorname{curl}_z \mathbf{A} = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \tag{40}$$

We can now write (39) as:

$$\int\limits_{(l)} A_s \, ds = \iint\limits_{(S)} [\operatorname{curl}_x \mathbf{A} \cos(n, X) + \operatorname{curl}_y \mathbf{A} \cos(n, Y) + \operatorname{curl}_z \mathbf{A} \cos(n, Z)] dS$$

or

$$\int\limits_{(l)} A_s \, ds = \iint\limits_{(S)} \operatorname{curl}_n \mathbf{A} \, dS, \tag{41}$$

where $\operatorname{curl}_n \mathbf{A}$ is the component of curl $\mathbf{A}$ along the normal (n) to (S). The line integral on the left is usually called the *circulation of vector* $\mathbf{A}$ *round the contour* (l) and Stokes' formula can be stated thus: *the circulation of the field round the contour of a given surface is equal to the integral over the surface of the normal component of the curl,* i.e. *equal to the flux of the curl through the surface.* We can define curl independently of the choice of axes with the aid of (41). Let (m) be a given direction through the point M, and let (σ) be a small plane surface through M normal to (m). We apply (41) to this area and use the mean value theorem:

$$\int\limits_{(\lambda)} A_s \, ds = \operatorname{curl}_m \mathbf{A}|_{M_1} \cdot \sigma, \quad \text{i.e.} \quad \operatorname{curl}_m \mathbf{A}|_{M_1} = \frac{\int\limits_{(\lambda)} A_s \, ds}{\sigma},$$

where (λ) is the contour of (σ) and M_1 is a point of the area. On indefinite contraction of the area to the point M, we get in the limit, as in the case of divergence, the value of the component of curl in any given direction (m) at the point M:

$$\operatorname{curl}_m \mathbf{A} = \lim_{(\sigma) \to M} \frac{\int\limits_{(\lambda)} A_s \, ds}{\sigma}. \tag{42}$$

Numerous examples will occur later of applications of curl and divergence, when the physical meaning of these will be explained.

110. Lamellar and solenoidal fields. We obtained in [108] the vector field grad $U(M)$ as the gradient of a scalar function $U(M)$. Such a field is called a *lamellar field*. Of course, not every vector field is a lamellar

field, and we now give the necessary and sufficient conditions for the field to be lamellar. The relationship $\mathbf{A} = \text{grad } U(M)$ is equivalent to [108]

$$A_x = \frac{\partial U}{\partial x}; \quad A_y = \frac{\partial U}{\partial y}; \quad A_z = \frac{\partial U}{\partial z},$$

i.e. to the fact that the expression

$$A_x \,\mathrm{d}x + A_y \,\mathrm{d}y + A_z \,\mathrm{d}z \tag{43}$$

is the total differential of a function. We saw in [73] that there are three necessary and sufficient conditions for this:

$$\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0; \ \frac{\partial A_x}{\partial z} - \frac{\partial Az}{\partial x} = 0; \ \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = 0,$$

which are in turn equivalent to the vanishing of the curl of the field: curl $\mathbf{A} = 0$, i.e. *the necessary and sufficient condition for a vector field to be lamellar is that its curl vanishes.* If this condition is satisfied, the potential is defined by [73] as a line integral:

$$U(M) = \int_{(M_0)}^{(M)} A_x \,\mathrm{d}x + A_y \,\mathrm{d}y + A_z \,\mathrm{d}z = \int_{(M_0)}^{(M)} A_s \,\mathrm{d}s. \tag{44}$$

With this, $\mathbf{A} = \text{grad } U(M)$, and [73]

$$\int_{(A)}^{(B)} A_s \,\mathrm{d}s = \int_{(A)}^{(B)} \text{grad}_s U(M) \,\mathrm{d}s = U(B) - U(A).$$

We can suppose that expression (43), whilst not itself a total differential, has an integrating factor, i.e. there exist the function $\mu(M)$ of the point M such that

$$\mu (A_x \,\mathrm{d}x + A_y \,\mathrm{d}y + A_z \,\mathrm{d}z) = \mathrm{d}U \tag{45}$$

is a total differential. We speak of a field of this sort as compound lamellar. We saw in [76] that the characteristic of such a field is the existence of the family of surfaces $U(M) = C$ orthogonal to a vector line of the field, whilst it follows from (45) that $\mu\mathbf{A} = \text{grad } U$ or

$$\mathbf{A} = \frac{1}{\mu} \text{grad } U,$$

i.e. the field $\mathbf{A}$ in this case differs from the lamellar field by a numerical factor $1/\mu$, with different values at different points of space.

The necessary and sufficient condition for a compound lamellar field is given by [76]:

$$A_x\left(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z}\right)+A_y\left(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x}\right)+A_z\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)=0,$$

which can be written as

$$\mathbf{A}\cdot\operatorname{curl}\mathbf{A}=0, \tag{46}$$

i.e. *the necessary and sufficient condition for the existence of a family of orthogonal surfaces to a vector line of the field is given by* (46), i.e. *the vectors* **A** *and* curl **A** *are perpendicular or else* curl **A** *vanishes.*

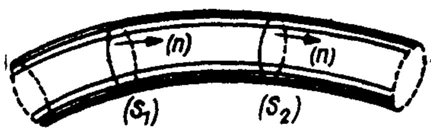

FIG. 95

If the space occupied by the field is multiply connected, the field potential defined by (44) may be a many-valued function.

We investigated above a vector field, the curl of which is zero, and discovered that the field is lamellar. A field for which the divergence is zero, i.e. where we have identically div $\mathbf{A}=0$, is called *solenoidal*. We have by (37) for this sort of field:

$$\iint_{(S)} A_n\,\mathrm{d}S=0, \tag{47}$$

where (S) is any closed surface inside which the field exists everywhere.

Let (S) be the part of a vector tube between its two sections (S_1) and (S_2) (Fig. 95). Since **A** lies in the tangent plane to the lateral part of (S), we have here $A_n=0$. If we take the normal (n) for both the sections (S_1) and (S_2) in the direction of movement along the tube, we shall get the inward normal in the case of (S_1) and the outward normal for (S_2) with respect to the piece of tube (S). We have on using (47):

$$\iint_{(S_2)} A_n\,\mathrm{d}S-\iint_{(S_1)} A_n\,\mathrm{d}S=0,$$

where the integral over (S_1) has the $(-)$ sign since (n) is in the opposite direction to the outward normal. This equation shows that

$$\iint_{(S)} A_n\,\mathrm{d}S \tag{48}$$

has the same value for every section (S) of a vector tube in the case of a solenoidal field. The integral represents the flux of the field through the section (S) and is usually called *the strength of the vector tube at section* (S). Hence we can say that the *strength is the same at all sections*

of a vector tube in the case of a solenoidal field. If the cross-sectional area increases on moving along a tube, i.e. the tube expands, the value of the flux intensity $\mathbf{A}_n$ generally speaking diminishes, so that the value of integral (48) remains unchanged.

111. Directed elementary areas. We now bring in directed elementary areas $\mathrm{d}\mathbf{S}$, corresponding to the directed elementary arcs of [109]. We suppose that two sides have been distinguished of a given surface, such that there are two opposite normal directions at any point of the surface, depending on which side is taken, and where continuous variation of either one direction is implied by continuous motion over the surface [64]. If the surface is closed, we get the inward and outward normals with respect to the volume enclosed by the surface. *We define a directed elementary surface* $\mathrm{d}\mathbf{S}$ *as the vector of length equal to the area* $\mathrm{d}S$ *of the elementary surface and with direction along the normal* (n) *to the element.* We agree to take (n) as the outward normal in the case of a closed surface, where the inward normal will be written (n_1).

The projections of $\mathrm{d}\mathbf{S}$ on the axes give us the projections of the elementary area on the corresponding coordinate planes, with the sign plus or minus, depending on whether the angle between (n) and the axis concerned is acute or obtuse.

Let $f(M)$ and $\mathbf{A}(M)$ be scalar and vector functions defined on the surface (S). We take the expressions:

$$\iint_{(S)} f(M)\,\mathrm{d}\mathbf{S} \tag{49}$$

$$\iint_{(S)} \mathbf{A}(M)\cdot\mathrm{d}\mathbf{S} \tag{49_1}$$

$$\iint_{(S)} \mathbf{A}(M)\times\mathrm{d}\mathbf{S}. \tag{49_2}$$

The first is a vector with the components:

$$\iint_{(S)} f(M)\cos(n, X)\,\mathrm{d}S; \qquad \iint_{(S)} f(M)\cos(n, Y)\,\mathrm{d}S;$$

$$\iint_{(S)} f(M)\cos(n, Z)\,\mathrm{d}S.$$

Expression (49_1) is a scalar:

$$\iint_{(S)} \mathbf{A}\cdot\mathrm{d}\mathbf{S} = \iint_{(S)} A_n\,\mathrm{d}S$$

whilst finally, (49_2) is a vector with components:

$$\iint_{(S)} [A_y \cos(n, Z) - A_z \cos(n, Y)]\, \mathrm{d}S,$$

$$\iint_{(S)} [A_z \cos(n, X) - A_x \cos(n, Z)]\, \mathrm{d}S,$$

$$\iint_{(S)} [A_x \cos(n, Y) - A_y \cos(n, X)]\, \mathrm{d}S.$$

Let (S) be closed and let (v) be the enclosed volume, $f(M)$ and $\mathbf{A}(M)$ being defined throughout the volume. By applying Ostrogradskii's formula, we can easily verify the following three equations:

$$\iint_{(S)} f \mathrm{d}\mathbf{S} = \iiint_{v} \operatorname{grad} f\, \mathrm{d}v \tag{50}$$

$$\iint_{(S)} \mathbf{A} \cdot \mathrm{d}\mathbf{S} = \iiint_{v} \operatorname{div} \mathbf{A}\, \mathrm{d}v \tag{50_1}$$

$$\iint_{(S)} \mathbf{A} \times \mathrm{d}\mathbf{S} = -\iiint_{v} \operatorname{curl} \mathbf{A}\, \mathrm{d}v. \tag{50_2}$$

Equation (50_1) is the same as (37). Equation (50_2) is verified as follows. The components of the left and right-hand sides along the x axis are given by the integrals:

$$\iint_{(S)} \left[A_y \cos(n, Z) - A_z \cos(n, Y)\right] \mathrm{d}S\,; \; -\iiint_{v} \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) \mathrm{d}v,$$

which are easily seen to be of the same magnitude by transforming the double integral with the aid of Ostrogradskii's formula [63].

Similarly, by using Stokes' formula and a directed elementary area, we can write the following expressions:

$$\int_{(l)} f\, \mathrm{d}s = -\iint_{(S)} \operatorname{grad} f \times \mathrm{d}\mathbf{S} \tag{51}$$

$$\int_{(l)} \mathbf{A} \cdot \mathrm{d}s = \iint_{(S)} \operatorname{curl} \mathbf{A} \cdot \mathrm{d}\mathbf{S}. \tag{51_1}$$

Here, (S) is a given surface and (l) its contour. The second formula is the same as (41), since $\operatorname{curl} \mathbf{A} \cdot \mathrm{d}\mathbf{S} = \operatorname{curl}_n A \cdot \mathrm{d}S$ by definition of scalar product. The components along OX of the left and right-hand sides of (51) are:

$$\int_{(l)} f\, \mathrm{d}x\,; \; -\iint_{(S)} \left[\frac{\partial f}{\partial y} \cos(n, Z) - \frac{\partial f}{\partial z} \cos(n, Y)\right] \mathrm{d}S\,;$$

these expressions are easily shown to be equal using (22) of [70].

112. Some formulae of vector analysis. We prove certain relationships between the vector operators introduced. We saw in [110] that the curl of a lamellar field is zero:

$$\operatorname{curl}\operatorname{grad} U = 0. \tag{52}$$

A rotational field is easily seen to have zero divergence:

$$\operatorname{div}\operatorname{curl}\mathbf{A} = 0. \tag{53}$$

We have, in fact:

$$\operatorname{div}\operatorname{curl}\mathbf{A} = \frac{\partial}{\partial x}\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) = 0.$$

We next consider the divergence of a lamellar field:

$$\operatorname{div}\operatorname{grad} U = \frac{\partial}{\partial x}\operatorname{grad}_x U + \frac{\partial}{\partial y}\operatorname{grad}_y U + \frac{\partial}{\partial z}\operatorname{grad}_z U,$$

or

$$\operatorname{div}\operatorname{grad} U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}. \tag{54}$$

The differential operator

$$\Delta U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} \tag{55}$$

is called *Laplace's operator*. From the left-hand side of (54), it is clearly independent of the choice of axes. On applying (38) to the vector grad U, we get ΔU at the point M defined as:

$$\Delta U = \lim_{(r_1)\to M} \frac{\displaystyle\iint_{(S_1)} \frac{\partial U}{\partial n}\, dS}{v_1}. \tag{56}$$

We have defined ΔU for the case of a scalar U. The symbol $\Delta\mathbf{A}$, where $\mathbf{A}$ is a vector field, denotes the vector with components ΔA_x, ΔA_y, ΔA_z. We now prove the following formulae:

$$\operatorname{curl}\operatorname{curl}\mathbf{A} = \operatorname{grad}\operatorname{div}\mathbf{A} - \Delta\mathbf{A}, \tag{57}$$

$$\operatorname{div}(f\mathbf{A}) = f\operatorname{div}\mathbf{A} + \operatorname{grad} f\cdot\mathbf{A}, \tag{57_1}$$

$$\operatorname{div}\mathbf{A}\times\mathbf{B} = \mathbf{B}\cdot\operatorname{curl}\mathbf{A} - \mathbf{A}\cdot\operatorname{curl}\mathbf{B}, \tag{57_2}$$

$$\operatorname{curl} f\mathbf{A} = \operatorname{grad} f\times\mathbf{A} + f\operatorname{curl}\mathbf{A}, \tag{57_3}$$

$$\Delta(\varphi\psi) = \psi\,\Delta\varphi + \varphi\,\Delta\psi + 2\operatorname{grad}\varphi\cdot\operatorname{grad}\psi. \tag{57_4}$$

We shall only prove the first of these, leaving the remaining proofs to the reader. We take the component along OX of the vector on the left-hand side of (57) and show that it coincides with the component of the vector on the right:

$$\operatorname{curl}_x \operatorname{curl} \mathbf{A} = \frac{\partial}{\partial y} \operatorname{curl}_z \mathbf{A} - \frac{\partial}{\partial z} \operatorname{curl}_y \mathbf{A} = \\ = \frac{\partial}{\partial y}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) - \frac{\partial}{\partial z}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right),$$

whence, on removing the brackets and adding and subtracting $\partial^2 A_x/\partial x^2$:

$$\operatorname{curl}_x \operatorname{curl} \mathbf{A} = \frac{\partial}{\partial x}\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}\right) - \left(\frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_x}{\partial y^2} + \frac{\partial^2 A_x}{\partial z^2}\right) = \\ = \frac{\partial}{\partial x} \operatorname{div} \mathbf{A} - \Delta A_x,$$

which is what we wished to prove. It may be remarked here that the independence of $\Delta\mathbf{A}$ on the choice of axes follows from (57), since

$$\Delta\mathbf{A} = \operatorname{grad} \operatorname{div} \mathbf{A} - \operatorname{curl} \operatorname{curl} \mathbf{A}.$$

113. Motion of a rigid body and small deformations. We saw in [106] that, in the case of rotation of a rigid body about a point O, the velocity of any point was given by:

$$\mathbf{v} = \mathbf{o} \times \mathbf{r},$$

where $\mathbf{o}$ is the instantaneous angular velocity vector and $\mathbf{r}$ is the radius vector $\overrightarrow{OM}$.

We get the most general case of motion of a rigid body on giving it in addition a translation of velocity $\mathbf{v}_0$, its total velocity being now:

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{o} \times \mathbf{r}. \tag{58}$$

Conversely, given the velocity field $\mathbf{v}$, the angular velocity vector is found as follows. First of all we notice that the vectors $\mathbf{v}_0$ are the same at any given instant at all points of the body, so that they are independent of x, y, z. Hence we have by (40), curl $\mathbf{v}_0 = 0$.

Let p, q, r be the axial components of $\mathbf{o}$, and let O be the origin. The components of the vector product $\mathbf{o} \times \mathbf{r}$ will be [104]: $qz - ry$, $rx - pz$, $py - qx$, so that by (40), the components of curl $\mathbf{o} \times \mathbf{r}$ are $2p$, $2q$, $2r$, whence the angular velocity vector is given in terms of $\mathbf{v}$ by

$$\mathbf{o} = \frac{1}{2} \operatorname{curl} \mathbf{v}. \tag{59}$$

This is the reason for curl $\mathbf{v}$ being alternatively called rot $\mathbf{v}$, i.e. the rotation of the velocity vector.

If we multiply $\mathbf{v}$ by the small element of time dt, we get a vector $\mathbf{v}\, dt$ which gives the approximate displacement of a point of the body in time dt. We thus

get a vector field of displacements of points of a rigid body:

$$\mathbf{A} = \mathbf{v}\, dt.$$

On returning to (58) and assuming that translation is absent, i.e. that the point O is fixed, we get the following expression for the displacement vector:

$$\mathbf{A} = \boldsymbol{\omega}_1 \times \mathbf{r}, \tag{60}$$

where $\boldsymbol{\omega}_1 = \boldsymbol{\omega}\, dt$ is a small vector directed along the axis of rotation and equal in magnitude to the small angular displacement in time dt. Let the components of this vector be p_1, q_1, r_1, and let (x, y, z) be the coordinates of a variable point of the rigid body. The components of $\mathbf{A}$ will be:

$$A_x = q_1 z - r_1 y; \quad A_y = r_1 x - p_1 z; \quad A_z = p_1 y - q_1 x.$$

Hence, as above, the vector of small rotations is easily found in terms of the displacement vector in the form:

$$\boldsymbol{\omega}_1 = \frac{1}{2} \operatorname{curl} \mathbf{A}. \tag{61}$$

Furthermore, the previous formulae show that the components of $\mathbf{A}$ are linear homogeneous functions of the coordinates (x, y, z).

We now consider the general case of a linear homogeneous deformation in which the components of the displacement vector are linear homogeneous functions of the coordinates:

$$\left.\begin{aligned} A_x &= a_1 x + b_1 y + c_1 z \\ A_y &= a_2 x + b_2 y + c_2 z \\ A_z &= a_3 x + b_3 y + c_3 z. \end{aligned}\right\} \tag{62}$$

We shall assume that coefficients a, b, and c are small, and shall confine ourselves to the case of a small volume (v) near the origin. Every point of the volume will be displaced by the vector $\mathbf{A}$ and its new coordinates will be:

$$\xi = x + A_x, \quad \eta = y + A_y; \quad \zeta = z + A_z,$$

i.e.

$$\left.\begin{aligned} \xi &= (1 + a_1) x + b_1 y + c_1 z \\ \eta &= a_2 x + (1 + b_2) y + c_2 z \\ \zeta &= a_3 x + b_3 y + (1 + c_3) z. \end{aligned}\right\} \tag{63}$$

This transformation will amount to a rotation of (v) as a rigid whole about O only in particular cases. In the general case, it will be accompanied by deformation of the volume, i.e. with a change in the distances between its points. We shall explain this in more detail.

The components of the curl of the displacement vector $\mathbf{A}$ are by (62): $b_3 - c_2$, $c_1 - a_3$, $a_2 - b_1$. If the transformation reduced to rotation of the elementary volume as a whole, we should have a displacement vector $\mathbf{A}^{(1)}$ with components

$$A_x^{(1)} = \frac{1}{2}(c_1 - a_3)\, z - \frac{1}{2}(a_2 - b_1)\, y; \quad A_y^{(1)} = \frac{1}{2}(a_2 - b_1)\, x - \frac{1}{2}(b_3 - c_2)\, z;$$

$$A_z^{(1)} = \frac{1}{2}(b_3 - c_2)\, y - \frac{1}{2}(c_1 - a_3)\, x.$$

On subtracting this vector from **A**, we can represent the latter as

$$\mathbf{A} = \mathbf{A}^{(1)} + \mathbf{A}^{(2)}, \tag{64}$$

where the vector of pure deformation $\mathbf{A}^{(2)}$ has the components:

$$\left.\begin{aligned} A_x^{(2)} &= a_1 x + \frac{1}{2}(b_1 + a_2)\,y + \frac{1}{2}(c_1 + a_3)\,z \\ A_y^{(2)} &= \frac{1}{2}(b_1 + a_2)\,x + b_2 y + \frac{1}{2}(c_2 + b_3)\,z \\ A_z^{(2)} &= \frac{1}{2}(c_1 + a_3)\,x + \frac{1}{2}(c_2 + b_3)\,y + c_3 z. \end{aligned}\right\} \tag{65}$$

It is easily seen that this latter is a potential vector, and in fact:

$$\mathbf{A}^{(2)} = \frac{1}{2}\,\mathrm{grad}\left[a_2 x^2 + b_2 y^2 + c_3 z^2 + (b_1 + a_2)\,xy + (c_1 + a_3)\,xz + (c_2 + b_3)\,yz\right],$$

whilst its curl is evidently zero.

We now find the change in an elementary volume as a result of deformation. The new volume after deformation will be given by:

$$v_1 = \iiint\limits_{(v)} \mathrm{d}\xi\,\mathrm{d}\eta\,\mathrm{d}\zeta.$$

If we carry out the change of variables in accordance with the formula of [60], we must write:

$$\mathrm{d}\xi\,\mathrm{d}\eta\,\mathrm{d}\zeta = \{(1 + a_1)\left[(1 + b_2)(1 + c_3) - c_2 b_3\right] + b_1\left[c_2 a_3 - a_2(1 + c_3)\right] + \\ + c_1\left[a_2 b_3 - (1 + b_2)\,a_3\right]\}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

On removing the brackets and keeping only the free terms and first powers of the small a, b, c, we get:

$$\mathrm{d}\xi\,\mathrm{d}\eta\,\mathrm{d}\zeta = \left[1 + (a_1 + b_2 + c_3)\right]\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z,$$

and the above formula gives:

$$v_1 = \iiint\limits_{(v)}\left[1 + (a_1 + b_2 + c_3)\right]\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = v + (a_1 + b_2 + c_3)\,v,$$

where v is the volume before deformation. The dilatation is

$$\frac{v_1 - v}{v} = a_1 + b_2 + c_3,$$

whilst the sum on the right is easily seen, by (62), to be div **A**, i.e. *the divergence of a displacement field gives the dilatation.*

114. Equation of continuity. Let **v** denote the velocity of flow of a fluid, and let us find the quantity of fluid flowing through a given surface (S) (Fig. 96), a small element of which is $\mathrm{d}S$. The particles occupying $\mathrm{d}S$ at time t move by an amount $\mathbf{v}\,\mathrm{d}t$ in time $\mathrm{d}t$; thus the quantity $\mathrm{d}Q$ of fluid which passes through $\mathrm{d}S$ in time $\mathrm{d}t$ occupies a cylinder of base $\mathrm{d}S$ and generator $\mathbf{v}\,\mathrm{d}t$. The height of the cylinder is

clearly $v_n\, dt$, where v_n is the projection of $\mathbf{v}$ on to the normal (n) to the surface, so that

$$dQ = \varrho v_n \, dt \, dS,$$

where ϱ is the density of the fluid. A negative dQ will be obtained if the angle (n, v) is obtuse. In the case of a closed surface, (n) is the direction of the outward normal, and a negative dQ is obtained if fluid flows into the volume enclosed by the surface through dS. The total quantity of fluid flowing through the surface per unit time will be:

$$Q = \iint\limits_{(S)} \varrho v_n \, dS, \tag{66}$$

where fluid flowing in is understood to take the minus sign.

The quantity of fluid occupying the volume (v) bounded by (S) is given by

$$\iiint\limits_{(v)} \varrho \, dv,$$

and this changes in time dt by an amount

$$dt \iiint\limits_{(v)} \frac{\partial \varrho}{\partial t} \, dv,$$

so that the increment in the quantity per unit time is

$$\iiint\limits_{(v)} \frac{\partial \varrho}{\partial t} \, dv,$$

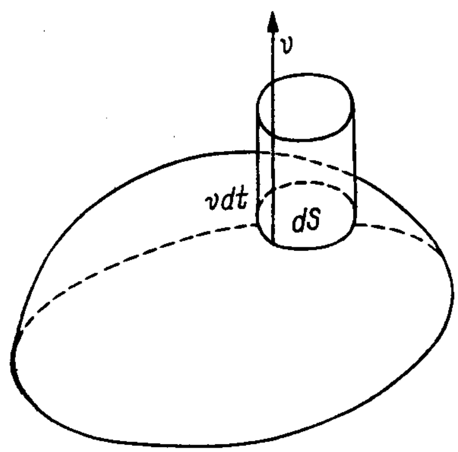

Fig. 96

whilst the quantity flowing out is given by the same integral with the reverse sign; hence we obtain two expressions for Q:

$$Q = \iint\limits_{(S)} \varrho v_n \, dS = -\iiint\limits_{(v)} \frac{\partial \varrho}{\partial t} \, dv,$$

or, on using (37):

$$Q = \iiint\limits_{(v)} \operatorname{div} (\varrho \mathbf{v}) \, dv = -\iiint\limits_{(v)} \frac{\partial \varrho}{\partial t} \, dv,$$

where the density ϱ is kept under the divergence sign since it may be variable, i.e. depend on the position of the point. The last expressions give us an equation which is valid for any volume inside a fluid:

$$\iiint\limits_{(v)} \left[\frac{\partial \varrho}{\partial t} + \operatorname{div} (\varrho \mathbf{v})\right] dv = 0.$$

It follows that the integrand must be *identically* zero†, and we get:

$$\frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho \mathbf{v}) = 0. \tag{67}$$

This important relationship between the density and velocity of flow of any fluid, compressible or not, is known as the equation of continuity. It may be written in an alternative form by considering the change in density of the fluid particle occupying the position (x, y, z) at the instant t.

We take $\varrho(t, x, y, z)$ as the fluid density at the point (x, y, z) at time t, and consider the change in density of a particle. As the particle moves, its density depends on t both directly and via (x, y, z), since its motion implies alteration of its coordinates. The total derivative of ϱ with respect to t is:

$$\frac{d\varrho}{dt} = \frac{\partial \varrho}{\partial t} + \frac{\partial \varrho}{\partial x}\frac{dx}{dt} + \frac{\partial \varrho}{\partial y}\frac{dy}{dt} + \frac{\partial \varrho}{\partial z}\frac{dz}{dt},$$

which can alternatively be written as

$$\frac{d\varrho}{dt} = \frac{\partial \varrho}{\partial t} + \frac{\partial \varrho}{\partial x} v_x + \frac{\partial \varrho}{\partial y} v_y + \frac{\partial \varrho}{\partial z} v_z,$$

or

$$\frac{d\varrho}{dt} = \frac{\partial \varrho}{\partial t} + \operatorname{grad} \varrho \cdot \mathbf{v}. \tag{68}$$

We can use (57_1) to write (67) in the form

$$\frac{\partial \varrho}{\partial t} + \operatorname{grad} \varrho \cdot \mathbf{v} + \varrho \operatorname{div} \mathbf{v} = 0,$$

i.e. by (68):

$$\frac{d\varrho}{dt} + \varrho \operatorname{div} \mathbf{v} = 0, \tag{69}$$

whence

$$\operatorname{div} \mathbf{v} = -\frac{1}{\varrho}\frac{d\varrho}{dt}.$$

The divergence of a velocity field $\mathbf{v}$ *thus gives the relative change in density per unit time of a fluid element situated at a given point.*

If the fluid is incompressible, this change must be zero, and (69) gives us the condition for incompressibility:

$$\operatorname{div} \mathbf{v} = 0. \tag{70}$$

† We proved in [71] that, if a double integral over any domain is zero, the integrand must be identically zero. The same proof may be used for triple integrals.

We deduced the continuity condition by finding two different expressions for the quantity of fluid flowing out of a volume. It is naturally assumed here that there are no sources in the volume, either positive or negative (sinks).

If the flow is irrotational or, in other words, potential, i.e. $\mathbf{v}$ is the lamellar vector:

$$\mathbf{v} = \operatorname{grad} \varphi,$$

we call φ the velocity potential. We get on substituting in (70):

$$\operatorname{div} \operatorname{grad} \varphi = 0, \qquad \text{i.e.} \qquad \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} = 0\,, \tag{71}$$

i.e. *the velocity potential must satisfy Laplace's equation* (71) *in the case of an incompressible fluid.*

115. Hydrodynamical equations for an ideal fluid. We shall understand by ideal fluid a deformable continuous medium such that the internal forces, whether the medium is in equilibrium or in motion, reduce to a normal pressure: so that if we distinguish a volume (v) of the fluid bounded by a surface (S), the action of the remainder of the fluid amounts to a force directed along the inward normal at every point of (S). We let p denote the magnitude of this force (pressure) per unit area. The pressure $p(M)$ at any given instant gives us a scalar field. The resultant of the pressure on the surface of (v) at a given instant gives us a scalar field. The resultant of the pressure in the surface of (v) is given by using (50):

$$-\iint\limits_{(S)} p\, \mathrm{d}\mathbf{S} = -\iiint\limits_{(v)} \operatorname{grad} p\, \mathrm{d}v,$$

where the $(-)$ sign is taken because a positive pressure acts along the inward normal, the vector $\mathrm{d}S$ being along the outward normal by hypothesis.

By d'Alembert's principle, the pressure must be in equilibrium with the external forces, which we denote by $\mathbf{F}$ per unit mass and which yield the resultant for volume (v):

$$\iiint\limits_{(v)} \varrho\, \mathbf{F}\, \mathrm{d}v,$$

to which must be added the inertia force on an elementary mass $-\varrho \mathrm{d}v\, \mathbf{W}$, where ϱ is the density and $\mathbf{W}$ is the acceleration of the fluid particle, so that the inertia force for (v) becomes

$$-\iiint\limits_{(v)} \varrho \mathbf{W}\, \mathrm{d}v.$$

d'Alembert's principle thus gives us:

$$\iiint\limits_{(v)} [\varrho \mathbf{F} - \operatorname{grad} p - \varrho \mathbf{W}]\, \mathrm{d}v = 0,$$

whence we can conclude, by the arbitrariness of (v) as above, that the integrand is zero, i.e.

$$\varrho \mathbf{W} = \varrho \mathbf{F} - \operatorname{grad} p. \tag{72}$$

This expression embraces the three fundamental equations of the hydrodynamics of an ideal fluid.

Let u, v, w be the components of the velocity, expressed as functions of the coordinates (x, y, z) of a point and time t. The component of the acceleration $\mathbf{W}$ along OX will be equal to the total differential of the component $u(x, y, z, t)$ of the velocity with respect to time, so that we can write:

$$W_x = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial t},$$

or

$$W_x = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} u + \frac{\partial u}{\partial y} v + \frac{\partial u}{\partial z} w.$$

Similarly:

$$W_y = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} u + \frac{\partial v}{\partial y} v + \frac{\partial v}{\partial z} w$$

$$W_z = \frac{\partial w}{\partial t} + \frac{\partial w}{\partial x} u + \frac{\partial w}{\partial y} v + \frac{\partial w}{\partial z} w.$$

Vector equation (72) thus leads to the three equations:

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} u + \frac{\partial u}{\partial y} v + \frac{\partial u}{\partial z} w &= F_x - \frac{1}{\varrho}\frac{\partial p}{\partial x} \\ \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} u + \frac{\partial v}{\partial y} v + \frac{\partial v}{\partial z} w &= F_y - \frac{1}{\varrho}\frac{\partial p}{\partial y} \\ \frac{\partial w}{\partial t} + \frac{\partial w}{\partial x} u + \frac{\partial w}{\partial y} v + \frac{\partial w}{\partial z} w &= F_z - \frac{1}{\varrho}\frac{\partial p}{\partial z}. \end{aligned} \tag{73}$$

These are referred to as the hydrodynamical equations in Euler's form. With them must be associated the continuity equation deduced in the previous article. Using the present notation, we can re-write (69) in the form:

$$\frac{\partial \varrho}{\partial t} + \frac{\partial \varrho}{\partial x} u + \frac{\partial \varrho}{\partial y} v + \frac{\partial \varrho}{\partial z} w + \varrho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) = 0. \tag{74}$$

A characteristic feature of the equations written is the fact that we have chosen the coordinates (x, y, z) of a point of space and time t as the independent variables when investigating the motion. In certain cases, the coordinates of the position of a fluid particle at the initial instant are chosen as independent variables instead of the coordinates (x, y, z) of a point; the hydrodynamical equations naturally have a different aspect with this choice.

116. Equations of sound propagation. Equations (72) and (73) are valid for gases as well as for fluids in the narrow sense of the word. The essential requirement is simply the hypothesis that the internal forces amount to a pressure only. We shall assume the motion sufficiently small for the terms containing products of the velocities and their derivatives with respect to the coordinates to be neglected on the left-hand sides of equations (73). With this, (73) become:

$$\frac{\partial u}{\partial t} = F_x - \frac{1}{\varrho}\frac{\partial p}{\partial x}; \quad \frac{\partial v}{\partial t} = F_y - \frac{1}{\varrho}\frac{\partial p}{\partial y}; \quad \frac{\partial w}{\partial t} = F_z - \frac{1}{\varrho}\frac{\partial p}{\partial z}, \tag{75}$$

or in vector form:

$$\frac{\partial \mathbf{v}}{\partial t} = \mathbf{F} - \frac{1}{\varrho} \operatorname{grad} p \,. \tag{76}$$

Similarly, neglecting the products of the velocity components and the derivatives of the density with respect to the coordinates in (74), we get:

$$\frac{\partial \varrho}{\partial t} + \varrho \operatorname{div} \mathbf{v} = 0. \tag{77}$$

Let ϱ_0 be the constant density of the medium in the state of rest. Let s be a small quantity characterizing the relative change in density during the motion and defined by the equation:

$$\varrho = \varrho_0 (1 + s).$$

From this:

$$\frac{d\varrho}{\varrho} = \frac{ds}{1+s} \sim ds,$$

the small quantity s being neglected in the denominator $(1 + s)$. It follows that we can take $\partial s/\partial t = (1/\varrho)\, \partial \varrho/\partial t$, and (77) gives:

$$\frac{\partial s}{\partial t} = -\operatorname{div} \mathbf{v}. \tag{78}$$

It can be assumed that the gradient of the pressure is proportional to the gradient of s, characterizing the compression or expansion, i.e.

$$\operatorname{grad} p = e \operatorname{grad} s,$$

where e is the modulus of elasticity of the medium. On substituting this in (76) and taking $\varrho = \varrho_0$ here, we get:

$$\frac{\partial v}{\partial t} = \mathbf{F} - \frac{e}{\varrho_0} \operatorname{grad} s.$$

We take the divergence of both sides of this equation:

$$\frac{\partial}{\partial t} \operatorname{div} \mathbf{v} = \operatorname{div} \mathbf{F} - \frac{e}{\varrho_0} \operatorname{div} \operatorname{grad} s.$$

On taking into account (78), we can write this as:

$$\frac{\partial^2 s}{\partial t^2} = a^2 \Delta s - \operatorname{div} \mathbf{F} \qquad \left(a = \sqrt{\frac{e}{\varrho_0}}\right). \tag{79}$$

This is the equation that must be satisfied by the function s of time and of the coordinates of a point. The differentiation with respect to t was taken outside the divergence sign when taking the divergence of $\partial \mathbf{v}/\partial t$, which is permissible since the result of differentiation is independent of the order in which it is carried out.

If there are no external forces, (79) becomes:

$$\frac{\partial^2 s}{\partial t^2} = a^2 \Delta s \qquad \left(a = \sqrt{\frac{e}{\varrho_0}}\right). \tag{80}$$

This latter is generally known as the *wave equation*. If we bear in mind that s characterizes the amount of compression or expansion, we can say that in our case the equation gives the law for sound propagation. The sound sources are the parts of space where div $\mathbf{F}$ differs from zero.

117. Equation of thermal conduction. We saw in [108] that the quantity of heat passing in time $\mathrm{d}t$ through an elementary surface $\mathrm{d}S$ may be taken as

$$\mathrm{d}Q = k\,\mathrm{d}t\,\mathrm{d}S\left|\frac{\partial U}{\partial n}\right| = k\,\mathrm{d}t\,\mathrm{d}S\,|\mathrm{grad}_n\,U(M)|,$$

where k is the coefficient of internal thermal conduction, U is the temperature and (n) is the direction of the normal to $\mathrm{d}S$. Let (S) be a closed surface bounding the volume (v), and let us find the total quantity of heat passing through (S). We easily see that:

$$\mathrm{d}Q = -\,\mathrm{d}t\iint\limits_{(S)} k\,\mathrm{grad}_n\,U\,\mathrm{d}S. \tag{81}$$

Here, if the temperature decreases in the direction (n) of the outward normal, $\partial U/\partial n < 0$, and the corresponding elements of the integral are negative, whilst the situation is reversed with increasing temperature. Noting the $(-)$ sign on the right of (81) and that heat flows in the direction of decreasing temperature, we can say that Q is the quantity of heat passing out of (v) in time $\mathrm{d}t$. The heat flowing into (v) will be given by (81) with the $(-)$ sign.

The quantity of heat given out may be alternatively obtained by observing the change of temperature inside the volume. We consider the elementary volume $\mathrm{d}v$; if its temperature increases by $\mathrm{d}U$ in time $\mathrm{d}t$, the quantity of heat expended must be proportional to the temperature rise and to the mass of the element, i.e. the quantity of heat is

$$\gamma\,\mathrm{d}U\cdot\varrho\,\mathrm{d}v = \gamma\varrho\frac{\partial U}{\partial t}\,\mathrm{d}t\,\mathrm{d}v,$$

where ϱ is the density of the body and γ a coefficient of proportionality known as the specific heat of the body. The heat given out by the total volume is thus

$$\mathrm{d}Q = -\,\mathrm{d}t\iiint\limits_{(v)} \gamma\,\varrho\,\frac{\partial U}{\partial t}\,\mathrm{d}v,$$

where the $(-)$ sign indicates that heat given out is to be understood. and not heat acquired.

On equating the two expressions obtained for dQ and using (37) of [109], we have

$$\iiint_{(v)} \gamma\varrho \frac{\partial U}{\partial t}\,dv = \iiint_{(v)} \operatorname{div}(k \operatorname{grad} U)\,dv, \tag{82}$$

i.e. we must have, for an arbitrary volume,

$$\iiint_{(v)} \left[\gamma\varrho \frac{\partial U}{\partial t} - \operatorname{div}(k \operatorname{grad} U)\right] dv = 0,$$

whence we get the differential equation for thermal conduction:

$$\gamma\varrho \frac{\partial U}{\partial t} = \operatorname{div}(k \operatorname{grad} U) \tag{83}$$

or

$$\gamma\varrho \frac{\partial U}{\partial t} = \frac{\partial}{\partial x}\left(k \frac{\partial U}{\partial x}\right) + \frac{\partial}{\partial y}\left(k \frac{\partial U}{\partial y}\right) + \frac{\partial}{\partial z}\left(k \frac{\partial U}{\partial z}\right). \tag{83_1}$$

This equation must be satisfied at all interior points of the body. The temperature U depends on the coordinates of the point and on time.

If the body is homogeneous, γ, ϱ and k are constants, and (83) can be written

$$\frac{\partial U}{\partial t} = a^2 \Delta U \qquad \left(a = \sqrt{\frac{k}{\gamma\varrho}}\right) \tag{84}$$

or

$$\frac{\partial U}{\partial t} = a^2 \left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}\right). \tag{84_1}$$

If the thermal phenomenon is stationary, i.e. the temperature is independent of time t and depends only on (x, y, z), (84) takes the form

$$\Delta U = 0, \text{ i.e. } \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0. \tag{85}$$

We have thus arrived at Laplace's equation, already encountered in [87] and [114], for the temperature in a stationary process.

We assumed when deducing the thermal conduction equation (83), that no heat sources were present in the body; if this were not the case, we should have to replace (82) by another equation:

$$\iiint_{(v)} \gamma\varrho \frac{\partial U}{\partial t}\,dv = \iiint_{(v)} \operatorname{div}(k \operatorname{grad} U)\,dv + \iiint_{(v)} e\,dv,$$

where the last term on the right-hand side represents the quantity of heat originating in the volume (v) and reckoned per unit time.

The integrand $e(t, M)$ gives the strength of the heat sources continuously distributed throughout (v) and can depend on time as well as

on the position of the point M. Instead of the differential thermal conduction equation (83) we should obtain an equation of the form

$$\gamma\varrho\frac{\partial U}{\partial t} = \operatorname{div}(k \operatorname{grad} U) + e \tag{86}$$

or, in the case of a homogeneous body, we should have instead of (84):

$$\frac{\partial U}{\partial t} = a^2 \Delta U + \frac{1}{\gamma\varrho} e\,. \tag{87}$$

Equations (87) and (84) are analogous to (79) and (80) of [116]. The presence of heat sources in the thermal conduction equation is analogous to that of external forces or, more precisely, sound sources div $\mathbf{F}$ in the propagation equations. The differential equations, (79) or (87), are non-homogeneous in both cases for this same reason, that they contain free terms, div $\mathbf{F}$ or e, which have to be looked on as given functions, in addition to the required functions s or U. An essential difference between equations (80) and (84) should be noticed. The first contains the second derivative of the required function with respect to time, whilst the second contains the first derivative. This circumstance makes an essential difference as regards the integration of the equations.

118. Maxwell's equations. The following vectors are introduced in the study of electromagnetic fields: the electric intensity $\mathbf{E}$, the magnetic force $\mathbf{H}$, the total current $\mathbf{r}$, the electric displacement $\mathbf{D}$, the magnetic induction $\mathbf{B}$. The two basic laws of electrodynamics, which are generalizations of the Biot–Savart and Faraday laws, can be written in the form

$$\int\limits_{(l)} H_s \, \mathrm{d}s = \frac{1}{c}\iint\limits_{(S)} r_n \, \mathrm{d}S, \tag{88}$$

$$\int\limits_{(l)} E_s \, \mathrm{d}s = -\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}t}\iint\limits_{(S)} B_n \, \mathrm{d}S, \tag{89}$$

where c is the speed of light in vacuo.

The first equation connects the circulation of the magnetic force round the contour of a surface with the flux of total current through the surface. The second connects the circulation of the electric intensity with the derivative with respect to time of the flux of magnetic induction through the surface. In these equations, (l) is an arbitrary closed contour bounding a surface (S). Furthermore, $\mathbf{D}$ and $\mathbf{B}$ are related to $\mathbf{E}$ and $\mathbf{H}$ in a quiescent homogeneous medium by:

$$\mathbf{D} = \varepsilon\mathbf{E}; \quad \mathbf{B} = \mu\mathbf{H},$$

where ε and μ are constants, called respectively the dielectric constant and the permeability of the medium. The total current is made up of two terms, the conduction current and the displacement current:

$$\mathbf{r} = \lambda \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t},$$

where λ is the coefficient of conductance of the medium. Hence (88) and (89) take the final form:

$$\int_{(l)} H_s \,\mathrm{d}s = \frac{1}{c} \iint_{(S)} \left(\lambda E_n + \varepsilon \frac{\partial E_n}{\partial t}\right) \mathrm{d}S, \tag{90$_1$}$$

$$\int_{(l)} E_s \,\mathrm{d}s = -\frac{1}{c} \frac{\mathrm{d}}{\mathrm{d}t} \iint_{(S)} \mu H_n \,\mathrm{d}S. \tag{90$_2$}$$

The integrals on the left-hand sides can be transformed to surface integrals by Stokes' formula:

$$\iint_{(S)} \mathrm{curl}_n \mathbf{H} \,\mathrm{d}S \quad \text{and} \quad \iint_{(S)} \mathrm{curl}_n \mathbf{E} \,\mathrm{d}S,$$

so that the equations now become:

$$\iint_{(S)} \left[c \,\mathrm{curl}_n \mathbf{H} - \left(\lambda E_n + \varepsilon \frac{\partial E_n}{\partial t}\right)\right] \mathrm{d}S = 0$$

$$\iint_{(S)} \left[c \,\mathrm{curl}_n \mathbf{E} + \mu \frac{\partial H_n}{\partial t}\right] \mathrm{d}S = 0.$$

We have from these, by the arbitrariness of surface (S) and therefore of the direction (n) of the normal:

$$c \,\mathrm{curl}\, \mathbf{H} = \lambda \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t}, \tag{91$_1$}$$

$$c \,\mathrm{curl}\, \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}. \tag{91$_2$}$$

These latter represent the differential form of Maxwell's equations, and amount to a set of six differential equations relating the six components

$$E_x,\ E_y,\ E_z,\ H_x,\ H_y,\ H_z.$$

It follows at once from (91_1) and (91_2) that in the present case the vector:

$$\lambda \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \quad \text{and} \quad \frac{\partial \mathbf{H}}{\partial t}$$

are solenoidal, since their divergences are respectively

$$c \,\mathrm{div}\,\mathrm{curl}\, \mathbf{H} \quad \text{and} \quad c \,\mathrm{div}\,\mathrm{curl}\, \mathbf{E}$$

and therefore vanish [112].

It can further be shown that $\mathbf{E}$ and $\mathbf{H}$ are themselves solenoidal in a given region of space, provided they are solenoidal there at some initial instant.

Before proving this, we introduce the two quantities

$$\operatorname{div} \varepsilon\mathbf{E} = \varrho_e = \varrho; \quad \operatorname{div} \mu\mathbf{H} = \varrho_m, \tag{92}$$

known respectively as the electric and magnetic charge densities. It now follows from the equation

$$\operatorname{div}\left(\lambda E + \varepsilon \frac{\partial \mathbf{E}}{\partial t}\right) = \frac{\lambda}{\varepsilon} \operatorname{div} \varepsilon\mathbf{E} + \frac{\partial}{\partial t} \operatorname{div}(\varepsilon\mathbf{E}) = 0$$

that

$$\frac{\lambda}{\varepsilon}\varrho + \frac{\partial \varrho}{\partial t} = 0,$$

and integration of this first order equation gives us [4]:

$$\varrho = \varrho_0 e^{-\frac{\lambda}{\varepsilon}t},$$

where ϱ_0 is the value of ϱ at $t = 0$. Consequently, if we have $\varrho_0 = 0$ at an initial instant, i.e.

$$\operatorname{div} \mathbf{E}_0 = 0,$$

we shall have $\varrho = 0$ for any t, i.e.

$$\operatorname{div} \mathbf{E} = 0.$$

Similarly, it follows from (91_2) that

$$\operatorname{div} \frac{\partial \mathbf{H}}{\partial t} = \frac{\partial}{\partial t} \operatorname{div} \mathbf{H} = 0,$$

and if $\operatorname{div} \mathbf{H}_0 = 0$, $\operatorname{div} \mathbf{H} = 0$ for any t.

The latter equation is equivalent to the vanishing condition for the magnetic charge, which is usually permissible.

Other equations can be deduced from Maxwell's, in which **E** and **H** appear separately. We take the curl of both sides of (91_2) and get:

$$-c \operatorname{curl} \operatorname{curl} \mathbf{E} = \mu \frac{\partial \operatorname{curl} \mathbf{H}}{\partial t}$$

or, on using (57_1) and (91_1):

$$c(\Delta\mathbf{E} - \operatorname{grad} \operatorname{div} \mathbf{E}) = \frac{\mu}{c} \frac{\partial}{\partial t}\left(\varepsilon \frac{\partial \mathbf{E}}{\partial t} + \lambda\mathbf{E}\right),$$

whence finally

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{\lambda}{\varepsilon} \frac{\partial \mathbf{E}}{\partial t} = \frac{c^2}{\varepsilon\mu}(\Delta\mathbf{E} - \operatorname{grad} \operatorname{div} \mathbf{E}). \tag{93}$$

Exactly the same form of equation may be obtained for **H**.

In the absence of electric charges, i.e. when $\operatorname{div} \mathbf{E} = 0$, (93) becomes:

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{\lambda}{\varepsilon} \frac{\partial \mathbf{E}}{\partial t} = \frac{c^2}{\varepsilon\mu} \Delta\mathbf{E}. \tag{94}$$

This is generally known as the equation of telegraphy, since it was first obtained when investigating the propagation of current along cables. Finally, if we are

concerned with a perfect dielectric, i.e. a non-conducting medium, we have $\lambda = 0$ and (94) becomes:

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} = a^2 \Delta \mathbf{E} \qquad \left(a = \frac{c}{\sqrt{\varepsilon\mu}}\right). \tag{95}$$

We have already encountered an equation of this form in [116].

If the process is stationary, i.e. vectors $\mathbf{E}$ and $\mathbf{H}$ are independent of t, equation (91_2) gives curl $\mathbf{E} = 0$, so that $\mathbf{E}$ is a potential vector: $\mathbf{E} = \operatorname{grad} \varphi$, whilst the first of equations (92) gives:

$$\operatorname{div} \operatorname{grad} \varphi = \frac{\varrho}{\varepsilon} \quad \text{or} \quad \Delta\varphi = \frac{\varrho}{\varepsilon}. \tag{96}$$

In places where $\varrho = 0$, i.e. electric charge is absent, we get Laplace's equation $\Delta\varphi = 0$ for the potential φ.

119. Laplace's operator in orthogonal coordinates. We introduced general curvilinear coordinates in space in [60]. We now consider a particular case of these coordinates, when the elementary volume referred to in [60] is a rectangular parallelepiped. This case of orthogonal curvilinear coordinates is the most important and the most frequently encountered in applications.

Let us take three new variables q_1, q_2, q_3 instead of Cartesian coordinates x, y, z in space:

$$\varphi(x, y, z) = q_1; \quad \psi(x, y, z) = q_2; \quad \omega(x, y, z) = q_3 \tag{97}$$

or, in the form when solved with respect to x, y, z:

$$x = \varphi_1(q_1, q_2, q_3); \quad y = \psi_1(q_1, q_2, q_3); \quad z = \omega_1(q_1, q_2, q_3). \tag{98}$$

On assigning constant values A, B, C to the new variables q_1, q_2, q_3, we get three families of coordinate surfaces, the equations of these in the x, y, z coordinates being:

$$\varphi(x, y, z) = A\,(\mathrm{I}); \quad \psi(x, y, z) = B\,(\mathrm{II}); \quad \omega(x, y, z) = C\,(\mathrm{III}). \tag{99}$$

We take any two coordinate surfaces from different families, say from families (II) and (III); these will intersect in a curve, with the equation

$$\psi(x, y, z) = B; \quad \omega(x, y, z) = C,$$

where B and C are definite constants. Only q_1 varies along this curve, and we can refer to it as the *coordinate line* q_1. Similarly, we have *coordinate lines* q_2 and q_3.

We find the square of an element of length in the new coordinates:

$$ds^2 = dx^2 + dy^2 + dz^2 = \left(\frac{\partial \varphi_1}{\partial q_1} dq_1 + \frac{\partial \varphi_1}{\partial q_2} dq_2 + \frac{\partial \varphi_1}{\partial q_3} dq_3\right)^2 +$$

$$+ \left(\frac{\partial \psi_1}{\partial q_1} dq_1 + \frac{\partial \psi_1}{\partial q_2} dq_2 + \frac{\partial \psi_1}{\partial \psi_3} dq_3\right)^2 + \left(\frac{\partial \omega_1}{\partial q_1} dq_1 + \frac{\partial \omega_1}{\partial q_2} dq_2 + \frac{\partial \omega_1}{\partial q_3} dq_3\right)^2. \quad (100)$$

On removing the brackets, we get a homogeneous second degree polynomial in dq_1, dq_2, dq_3. Next we find the conditions for the polynomial not to contain product terms in different differentials dq.

We can take, say, the term in the product $dq_1\, dq_2$ in (100). The coefficient of the product will be:

$$2\left(\frac{\partial \varphi_1}{\partial q_1} \cdot \frac{\partial \varphi_1}{\partial q_2} + \frac{\partial \psi_1}{\partial q_1} \cdot \frac{\partial \psi_1}{\partial q_2} + \frac{\partial \omega_1}{\partial q_1} \cdot \frac{\partial \omega_1}{\partial q_2}\right). \quad (101)$$

An elementary volume in the new coordinates (Fig. 97) will be bounded by three pieces of coordinate surfaces. Three edges, AB, AC, AD, will leave the vertex A of its base, corresponding to the values q_1, q_2, q_3 of the new coordinates. Only q_1 varies along AB, only q_2 along AC, and only q_3 along AD. If we consider the first and second edges, functions (98) are functions of q_1 only on the first edge, and the direction-cosines of the tangent are proportional to [I, 160]

$$\frac{\partial \varphi_1}{\partial q_1}, \quad \frac{\partial \psi_1}{\partial q_1}, \quad \frac{\partial \omega_1}{\partial q_1}.$$

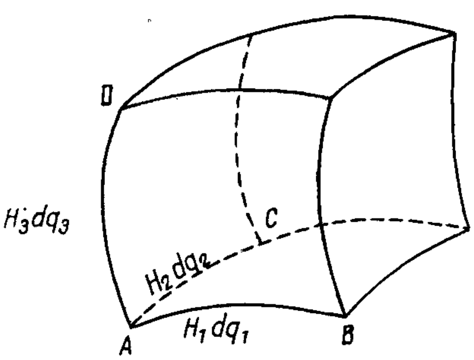

FIG. 97

Similarly, the direction-cosines of the tangent to the second edge are proportional to:

$$\frac{\partial \varphi_1}{\partial q_2}, \quad \frac{\partial \psi_1}{\partial q_2}, \quad \frac{\partial \omega_1}{\partial q_2}.$$

The vanishing of (101) is thus equivalent to these two edges being perpendicular. If we also require the vanishing of the coefficients of $dq_1\, dq_3$ and $dq_2\, dq_3$ in (100), this is equivalent to requiring that all three edges of the elementary volume in the new coordinates should be mutually perpendicular. Hence, *the necessary and sufficient condition or a system of curvilinear coordinates to be orthogonal is that the expression*

for ds^2 *should contain only terms in the squares of the differentials,* i.e. *only terms in* dq_1^2, dq_2^2, *and* dq_3^2.

We shall assume in future that the curvilinear coordinates are orthogonal.

We now have for ds^2 an expression of the form

$$ds^2 = H_1^2\,dq_1^2 + H_2^2\,dq_2^2 + H_3^2\,dq_3^2, \tag{102}$$

where

$$\left.\begin{aligned} H_1^2 &= \left(\frac{\partial\varphi_1}{\partial q_1}\right)^2 + \left(\frac{\partial\psi_1}{\partial q_1}\right)^2 + \left(\frac{\partial\omega_1}{\partial q_1}\right)^2 \\ H_2^2 &= \left(\frac{\partial\varphi_1}{\partial q_2}\right)^2 + \left(\frac{\partial\psi_1}{\partial q_2}\right)^2 + \left(\frac{\partial\omega_1}{\partial q_2}\right)^2 \\ H_3^2 &= \left(\frac{\partial\varphi_1}{\partial q_3}\right)^2 + \left(\frac{\partial\psi_1}{\partial q_3}\right)^2 + \left(\frac{\partial\omega_1}{\partial q_3}\right)^2. \end{aligned}\right\} \tag{103}$$

If we recall that only one variable changes along each of the edges of an elementary volume, we obtain for the lengths of the edges, by (102):

$$ds_1 = H_1\,dq_1; \quad ds_2 = H_2\,dq_2; \quad ds_3 = H_3\,dq_3, \tag{104}$$

and an elementary volume in the new coordinates is given by the expression

$$dv = ds_1\,ds_2\,ds_3 = H_1\,H_2\,H_3\,dq_1\,dq_2\,dq_3. \tag{105}$$

Now let $\mathbf{A}$ be a vector field in space; we know from [109] that its divergence at a point M is given by

$$\operatorname{div}\mathbf{A} = \lim_{(v_1)\to M} \frac{\iint\limits_{(S_1)} A_n\,dS}{v_1},$$

where (S_1) is the surface bounding a volume (v_1) of magnitude V_1 which contains the point M and contracts indefinitely towards it. We apply this to an elementary volume in curvilinear coordinates q_1, q_2, q_3, and find the flux of the field through the surface of the volume. We start by finding the flux through the right and left-hand faces. The coordinates have the values q_1, q_2, q_3 at the base vertex A, whilst q_1 will have to be replaced by $(q_1 + dq_1)$ on the right-hand face. Furthermore, the direction of the outward normal on the right-hand face coincides with the direction of the coordinate line q_1, the direction on the left-hand face being the reverse. The component A_n along the outward normal (n) is thus Aq_1 on the right-hand face and $(-Aq_1)$ on the left-hand face, where Aq_1 is the projection of vector $\mathbf{A}$ on the tangent to the coordinate line q_1 or, as we usually say, on the coordinate line q_1.

In view of the smallness of the faces, we can replace the surface integrals $\iint A_n \,\mathrm{d}S$ over them by the simple product of the integrand and the area of the corresponding face, which gives us for the flux through the right and left-hand faces:

$$A_{q_1} \mathrm{d}s_2 \,\mathrm{d}s_3 \,|_{q_1+dq_1} \text{ and } -A q_1 \,\mathrm{d}s_2 \,\mathrm{d}s_3 \,|_{q_1},$$

the flux through both faces being

$$A_{q_1} \mathrm{d}s_2 \,\mathrm{d}s_3 \,|_{q_1+dq_1} \qquad -A_{q_1} \mathrm{d}s_2 \,\mathrm{d}s_3 \,|_{q_1}$$

or, by (104):

$$A_{q_1} H_2 H_3 \,\mathrm{d}q_2 \,\mathrm{d}q_3 \,|_{q_1+dq_1} - A_{q_1} H_2 H_3 \,\mathrm{d}q_2 \,\mathrm{d}q_3 \,|_{q_1} =$$
$$= [H_2 H_3 A_{q_1} |_{q_1+dq_1} - H_2 H_3 A_{q_1} |_{q_1}] \,\mathrm{d}q_2 \,\mathrm{d}q_3 .$$

We finally replace the increment of the functions by the differential, and get for the flux through the right and left-hand faces:

$$\frac{\partial(H_2 H_3 A q_1)}{\partial q_1} \,\mathrm{d}q_1 \,\mathrm{d}q_2 \,\mathrm{d}q_3.$$

Similarly, the flux through the front and rear faces is

$$\frac{\partial(H_3 H_1 A q_2)}{\partial q_2} \,\mathrm{d}q_1 \,\mathrm{d}q_2 \,\mathrm{d}q_3$$

and through the upper and lower faces:

$$\frac{\partial(H_1 H_2 A q_3)}{\partial q_3} \,\mathrm{d}q_1 \,\mathrm{d}q_2 \,\mathrm{d}q_3.$$

We add these three expressions and divide by the volume obtained from (105) and arrive at an expression for the divergence of the field in orthogonal curvilinear coordinates:

$$\operatorname{div} \mathbf{A} = \frac{1}{H_1 H_2 H_3} \left[\frac{\partial(H_2 H_3 A q_1)}{\partial q_1} + \frac{\partial(H_3 H_1 A q_2)}{\partial q_2} + \frac{\partial(H_1 H_2 A q_3)}{\partial q_3} \right]. \qquad (106)$$

Now let $\mathbf{A}$ be a potential field, i.e. the field of the gradient of a function $U(M)$, so that $\mathbf{A} = U$.

The component A_{q_1} is now the derivative of U with respect to the direction q_1:

$$A_{q_1} = \lim_{\Delta s_1 \to 0} \frac{\Delta U}{\Delta s_1} = \frac{1}{H_1} \frac{\partial U}{\partial q_1},$$

and similarly:

$$A_{q_2} = \frac{1}{H_2} \frac{\partial U}{\partial q_2}; \qquad A_{q_3} = \frac{1}{H_3} \frac{\partial U}{\partial q_3}.$$

Substitution of these expressions in (106) gives us the expression for Laplace's operator in curvilinear orthogonal coordinates:

$$\Delta U = \operatorname{div} \operatorname{grad} U = \frac{1}{H_1 H_2 H_3} \left[\frac{\partial}{\partial q_1} \left(\frac{H_2 H_3}{H_1} \frac{\partial U}{\partial q_1} \right) + \right.$$

$$\left. + \frac{\partial}{\partial q_2} \left(\frac{H_3 H_1}{H_2} \frac{\partial U}{\partial q_2} \right) + \frac{\partial}{\partial q_3} \left(\frac{H_1 H_2}{H_3} \frac{\partial U}{\partial q_3} \right) \right]. \tag{107}$$

Laplace's equation $\Delta U = 0$ takes the form, in coordinates q_1, q_2, q_3:

$$\frac{\partial}{\partial q_1} \left(\frac{H_2 H_3}{H_1} \frac{\partial U}{\partial q_1} \right) + \frac{\partial}{\partial q_2} \left(\frac{H_3 H_1}{H_2} \frac{\partial U}{\partial q_2} \right) +$$

$$+ \frac{\partial}{\partial q_3} \left(\frac{H_1 H_2}{H_3} \frac{\partial U}{\partial q_3} \right) = 0. \tag{108}$$

1. Spherical coordinates. In this case, expressions (98) become [59]:

$$x = r \sin\theta \cos\varphi; \quad y = r \sin\theta \sin\varphi; \quad z = r\cos\theta,$$

where $q_1 = r$, $q_2 = \theta$ and $q_3 = \varphi$. We find ds^2:

$$ds^2 = (\sin\theta\cos\varphi\, dr + r\cos\theta\cos\varphi\, d\theta - r\sin\theta\sin\varphi\, d\varphi)^2 +$$

$$+ (\sin\theta\sin\varphi\, dr + r\cos\theta\sin\varphi\, d\theta + r\sin\theta\cos\varphi\, d\varphi)^2 +$$

$$+ (\cos\theta\, dr - r\sin\theta\, d\theta)^2,$$

or, on removing the brackets:

$$ds^2 = dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2, \tag{109}$$

i.e. $H_1 = 1$, $H_2 = r$, $H_3 = r\sin\theta$, with $0 \leqslant \theta \leqslant \pi$, so that $H_3 \geqslant 0$. Substitution in (108) gives us Laplace's equation in spherical coordinates:

$$\frac{\partial}{\partial r}\left(r^2 \sin\theta \frac{\partial U}{\partial r} \right) + \frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial U}{\partial \theta} \right) + \frac{\partial}{\partial \varphi}\left(\frac{1}{\sin\theta} \frac{\partial U}{\partial \varphi} \right) = 0$$

or

$$\frac{\partial}{\partial r}\left(r^2 \frac{\partial U}{\partial r} \right) + \frac{1}{\sin\theta} \frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial U}{\partial \theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2 U}{\partial \varphi^2} = 0. \tag{110}$$

We find the solution of this equation depending only on the radius vector. With this, we have to take $\partial U/\partial\theta = \partial U/\partial\varphi = 0$, so that

$$\frac{\partial}{\partial r}\left(r^2 \frac{\partial U}{\partial r} \right) = 0,$$

whence

$$r^2 \frac{\partial U}{\partial r} = -C_1 \quad \text{or} \quad \frac{\partial U}{\partial r} = -\frac{C_1}{r^2},$$

and we find on integration:

$$U = \frac{C_1}{r} + C_2, \tag{111}$$

where C_1 and C_2 are arbitrary constants. We recall that r is the distance of the variable point M to the fixed point M_0, which we can choose as the origin. In particular, with $C_1 = 1$ and $C_2 = 0$, we have the solution $1/r$, already mentioned in [87].

2. Cylindrical coordinates. Here,

$$x = \varrho \cos \varphi; \qquad y = \varrho \sin \varphi; \qquad z = z,$$

so that $q_1 = \varrho$, $q_2 = \varphi$, $q_3 = z$. We have for ds^2:

$$ds^2 = d\varrho^2 + \varrho^2 \, d\varphi^2 + dz^2,$$

whence $H_1 = 1$, $H_2 = \varrho$, $H_3 = 1$, and Laplace's equation becomes in cylindrical coordinates, by (108):

$$\frac{\partial}{\partial \varrho}\left(\varrho \frac{\partial U}{\partial \varrho} \right) + \frac{\partial}{\partial \varphi}\left(\frac{1}{\varrho} \frac{\partial U}{\partial \varphi} \right) + \frac{\partial}{\partial z}\left(\varrho \frac{\partial U}{\partial z} \right) = 0,$$

or

$$\frac{\partial}{\partial \varrho}\left(\varrho \frac{\partial U}{\partial \varrho} \right) + \frac{1}{\varrho} \frac{\partial^2 U}{\partial \varphi^2} + \varrho \frac{\partial^2 U}{\partial z^2} = 0. \tag{112}$$

It is easily shown, as above, that the solution depending only on the distance ϱ from OZ is:

$$U = C_1 \log \varrho + C_2, \tag{113}$$

Let U be independent of z, i.e. U has the same value at corresponding points of all the planes parallel to the XY plane. It is now sufficient to consider the values of U on the XY plane only (the plane case). Laplace's equation in rectilinear rectangular coordinates gives here:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0.$$

If the plane is referred to polar coordinates (ϱ, φ), we get the equation, by (112):

$$\frac{\partial}{\partial \varrho}\left(\varrho \frac{\partial U}{\partial \varrho} \right) + \frac{1}{\varrho} \frac{\partial^2 U}{\partial \varphi^2} = 0.$$

It is clear from (113) that in the plane case a solution of Laplace's equation will be given by $\log \varrho$, where ϱ is the distance of a variable point from a fixed point. Of course we could take the solution $\log 1/\varrho = -\log \varrho$ instead of $\log \varrho$. The fundamental solution of Laplace's

equation in three-dimensional space is thus the inverse of the distance of a variable point from a fixed point, whilst it becomes the logarithm of this inverse distance, or of the distance itself, in the plane case.

120. Differentiation in the case of a variable field. Let a scalar field $U(t, M)$ and a vector field $\mathbf{A}(t, M)$ be given in space, the field varying with time in both cases, i.e. the value of the scalar or vector is a function of time at every point. Further, let all space be given a motion represented by the velocity field $\mathbf{v}$. We shall suppose this latter vector to be also dependent on time t.

We investigate the variation of U with time. We can do this in two ways:

1. We can fix our attention on a definite point of space and find the rate of change of U at this point. We arrive at the partial derivative $\partial U/\partial t$, which may be called the *local derivative*, since it relates to a definite position in space.

2. We can find the rate of change of U whilst fixing our attention on a definite particle of the moving medium (substance). When differentiating with respect to time, we now have to take into account the motion of the points of the medium themselves, i.e. we have to differentiate U both directly with respect to t and indirectly via the coordinates (x, y, z) of the point M. We arrive in this case at the total, or as it is sometimes called, the *particle derivative:*

$$\frac{\mathrm{d}U}{\mathrm{d}t} = \frac{\partial U}{\partial t} + \frac{\partial U}{\partial x}\,\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial U}{\partial y}\,\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial U}{\partial z}\,\frac{\mathrm{d}z}{\mathrm{d}t} =$$

$$= \frac{\partial U}{\partial t} + \frac{\partial U}{\partial x}\,v_x + \frac{\partial U}{\partial y}\,v_y + \frac{\partial U}{\partial z}\,v_z\,,$$

which may be written in the condensed form:

$$\frac{\mathrm{d}U}{\mathrm{d}t} = \frac{\partial U}{\partial t} + \mathbf{v}\cdot \operatorname{grad} U\,. \tag{114}$$

We had an example of a particle derivative in [114], where we considered the total derivative with respect to time of the density of a particle of a continuous medium in motion.

Similarly, we have for the variable vector $\mathbf{A}(t, M)$ in a moving medium:

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = \frac{\partial \mathbf{A}}{\partial t} + \frac{\partial \mathbf{A}}{\partial x}\,v_x + \frac{\partial \mathbf{A}}{\partial y}\,v_y + \frac{\partial \mathbf{A}}{\partial z}\,v_z$$

or

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = \frac{\partial \mathbf{A}}{\partial t} + (\mathbf{v}\operatorname{grad})\,\mathbf{A}, \tag{115}$$

where the meaning of the symbol ($\mathbf{v}$ grad) is given by:

$$(\mathbf{v}\operatorname{grad}) = v_x\frac{\partial}{\partial x} + v_y\frac{\partial}{\partial y} + v_z\frac{\partial}{\partial z}\;.$$

The first term in (114) or (115), i.e. the partial derivative with respect to time, represents the change at a given place, whilst the second term represents the result of the motion of the medium itself.

We now establish certain formulae for differentiating integrals over domains related to moving media. The dependence of the integral on time is here due both to the integrand being a function of time and to the domain of integration varying with time. We can find the derivative with respect to time t by looking on the two-fold dependence on t as a dependence on two variables then applying the rule for differentiating functions of a function [I, 69]. It amounts in essence to a case of the principle of superposition of indefinitely small operations.

The derivative of the integral with respect to t will consist of two terms: the first is found by assuming that the domain of integration is fixed and is given by simple differentiation with respect to t under the integral sign [80], whilst the second only takes into account the change in the domain of integration, the integrand being meantime assumed constant.

We consider various cases.

1. Let (v) be a variable volume and $U(t, M)$ a scalar function. We find the expression for the derivative:

$$\frac{\mathrm{d}}{\mathrm{d}t}\iiint\limits_{(v)} U\,\mathrm{d}v.$$

Every element $\mathrm{d}S$ of the surface (S) bounding (v) describes a volume $\mathrm{d}t\, v_n\, \mathrm{d}S$ in time $\mathrm{d}t$, where (n) is the direction of the outward normal to (S) [114].

On multiplying this change in volume by the corresponding value of the integrand U and summing over the total surface (S), we get the change in the value of the integral due to the change in the domain (v) itself†:

$$\mathrm{d}t\iint\limits_{(S)} U v_n\,\mathrm{d}S.$$

On dividing by $\mathrm{d}t$ and adding the term due to the change in the integrand, we get the expression for the derivative of the integral in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\iiint\limits_{(v)} U\,\mathrm{d}v = \iiint\limits_{(v)} \frac{\partial U}{\partial t}\,\mathrm{d}v + \iint\limits_{(S)} U v_n\,\mathrm{d}S,$$

whence, on applying Ostrogradskii's formula, we have:

$$\frac{\mathrm{d}}{\mathrm{d}t}\iiint\limits_{(v)} U\,\mathrm{d}v = \iiint\limits_{(v)} \left[\frac{\partial U}{\partial t} + \operatorname{div}(U\mathbf{v})\right]\mathrm{d}v. \tag{116}$$

On expressing $\partial U/\partial t$ in terms of $\mathrm{d}U/\mathrm{d}t$ in accordance with (114) and using (57_1) [112],

$$\operatorname{div}(U\mathbf{v}) = U\operatorname{div}\mathbf{v} + \mathbf{v}\cdot\operatorname{grad} U,$$

we can re-write (116) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\iiint\limits_{(v)} U\,\mathrm{d}v = \iiint\limits_{(v)} \left[\frac{\mathrm{d}U}{\mathrm{d}t} + U\operatorname{div}\mathbf{v}\right]\mathrm{d}v. \tag{117}$$

† If the angle between the direction of $\mathbf{v}$ and (n) is obtuse, v_n will be negative and the change in volume $\mathrm{d}t\, v_n\, \mathrm{d}S$ will also be negative.

2. We now consider the derivative of the flux of a variable vector field $\mathbf{A}(t, M)$ through a moving surface (S):

$$\frac{\mathrm{d}}{\mathrm{d}t}\iint\limits_{(S)} A_n\,\mathrm{d}S.$$

Here, (S) is a surface related to a moving medium and (n) is the direction of the normal to (S). One term in the required expression will be

$$\iint\limits_{(S)} \frac{\partial A_n}{\partial t}\,\mathrm{d}S. \tag{118}$$

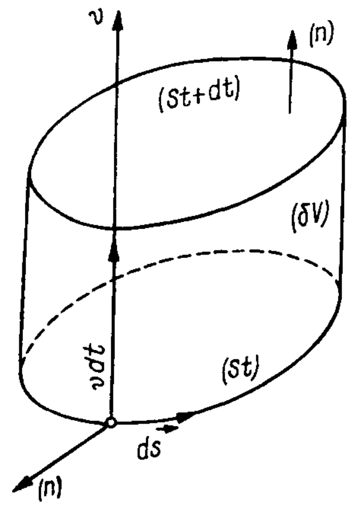

FIG. 98

The second term, due to the movement of (S) itself, is found as follows. Let (l) be the contour of the surface and $\mathrm{d}\mathbf{s}$ a directed element of the contour; we define the direction of (l) later (Fig. 98). Surface (S) describes a volume (δV) in time $\mathrm{d}t$, the three boundary surfaces being: the position (S_t) of (S) at time t, the position $(S_{t+\mathrm{d}t})$ at time $t+\mathrm{d}t$, and the surface (S') described by (l) in the interval $\mathrm{d}t$. An elementary area of (S') will be

$$\mathrm{d}S' = |\,\mathrm{d}\mathbf{s}\times\mathbf{v}\,|\,\mathrm{d}t.$$

Let (n) be the direction of the normal to (S_t) and $(S_{t+\mathrm{d}t})$, taken in the same direction, and let the normal to $(S_{t+\mathrm{d}t})$ be directed outwards from (δV). Let (n) also denote the direction of the outward normal from (δV) to (S'), and let the direction of (l) be specified such that $\mathrm{d}\mathbf{s}$, $\mathbf{v}$ and (n) to (S') have the same orientation as the axes. With this, clearly,

$$A_n\,\mathrm{d}S' = \mathbf{A}\cdot(\mathrm{d}\mathbf{s}\times\mathbf{v})\,\mathrm{d}t,$$

so that Ostrogradskii's formula gives us:

$$\iint\limits_{(S_{t+\mathrm{d}t})} A_n\,\mathrm{d}S - \iint\limits_{(S_t)} A_n\,\mathrm{d}S + \mathrm{d}t\int\limits_{(l)} \mathbf{A}\cdot(\mathrm{d}\mathbf{s}\times\mathbf{v}) = \iiint\limits_{(\delta V)} \operatorname{div}\mathbf{A}\,\mathrm{d}v. \tag{119}$$

The $(-)$ sign in front of the integral over (S_t) is due to the normal (n) over (S_t) being directed inwards to (δV). But we know from [105] that

$$\mathbf{A}\cdot(\mathrm{d}\mathbf{s}\times\mathbf{v}) = \mathrm{d}\mathbf{s}\cdot(\mathbf{v}\times\mathbf{A}) = (\mathbf{v}\times\mathbf{A})_s\,\mathrm{d}s,$$

where $(\mathbf{v}\times\mathbf{A})_s$ is the projection of $\mathbf{v}\times\mathbf{A}$ on the direction $\mathrm{d}\mathbf{s}$, and hence, by Stokes' formula:

$$\int\limits_{(l)} \mathbf{A}\cdot(\mathrm{d}\mathbf{s}\times\mathbf{v}) = \int\limits_{(l)} (\mathbf{v}\times\mathbf{A})_s\,\mathrm{d}s = \iint\limits_{(S_t)} \operatorname{curl}_n(\mathbf{v}\times\mathbf{A})\,\mathrm{d}S.$$

On dividing (δV) into elementary volumes $\mathrm{d}v = v_n\,\mathrm{d}S\,\mathrm{d}t$, where $\mathrm{d}S$ is an element of (S_t), we get by (119):

$$\iint\limits_{(S_{t+\mathrm{d}t})} A_n\,\mathrm{d}S - \iint\limits_{(S_t)} A_n\,\mathrm{d}S = \mathrm{d}t\iint\limits_{(S_t)} \left[v_n \operatorname{div}\mathbf{A} - \operatorname{curl}_n(\mathbf{v}\times\mathbf{A})\right]\mathrm{d}S.$$

We divide both sides by dt and pass to the limit to obtain the term of the derivative due to motion of (S). We add term (118) and get finally:

$$\frac{\mathrm{d}}{\mathrm{d}t}\iint\limits_{(S)} A_n\,\mathrm{d}S = \iint\limits_{(S)}\left[\frac{\partial A_n}{\partial t} + v_n \operatorname{div}\mathbf{A} + \operatorname{curl}_n(\mathbf{A}\times\mathbf{v})\right]\mathrm{d}S\,. \tag{120}$$

The term in $\operatorname{curl}_n(\mathbf{A}\times\mathbf{v})$ is absent if (S) is a closed surface, and the formula in this case follows at once from (116). Suppose, in fact, that (v) is a variable volume bounded by the closed surface (S); we obtain, on applying Ostrogradskii's formula and (116):

$$\frac{d}{\mathrm{d}t}\iint\limits_{(S)} A_n\,\mathrm{d}S = \frac{\mathrm{d}}{\mathrm{d}t}\iiint\limits_{(v)}\operatorname{div}\mathbf{A}\,\mathrm{d}v = \iiint\limits_{(v)}\left[\frac{\partial}{\mathrm{d}t}\operatorname{div}\mathbf{A} + \operatorname{div}(\mathbf{v}\operatorname{div}\mathbf{A})\right]\mathrm{d}v =$$

$$= \iiint\limits_{(v)}\operatorname{div}\left[\frac{\partial\mathbf{A}}{\partial t} + \mathbf{v}\operatorname{div}\mathbf{A}\right]\mathrm{d}v = \iint\limits_{(S)}\left(\frac{\partial\mathbf{A}_n}{\partial t} + v_n\operatorname{div}\mathbf{A}\right)\mathrm{d}S\,.$$

3. We now find the derivative of the circulation of a variable vector over a moving curve:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{(l)} A_s\,\mathrm{d}s\,.$$

As usual, one term in the required expression will be

$$\int\limits_{(l)}\frac{\partial A_s}{\partial s}\,\mathrm{d}s. \tag{121}$$

The additional term, due to the motion of the curve itself, is obtained as follows. The curve (l) describes a surface (δS) in time dt, the four boundary lines being (Fig. 99): curve A_1A_2, which is the position (l_t) of (l) at time t; curve B_1B_2, the position $(l_{t+\mathrm{d}t})$ of (l) at time $t+\mathrm{d}t$; and finally, the curves A_1B_1 and A_2B_2, described by the ends A_1 and A_2 of (l) in time dt. Stokes' formula gives:

$$\int\limits_{(l_t)} A_s\,\mathrm{d}s + \int\limits_{(A_2B_2)} A_s\,\mathrm{d}s - \int\limits_{(l_{t+\mathrm{d}t})} A_s\,\mathrm{d}s + \int\limits_{(B_1A_1)} A_s\,\mathrm{d}s = \iint\limits_{(\delta S)}\operatorname{curl}_n\mathbf{A}\,\mathrm{d}S, \tag{122}$$

the integrations over (l_t) and $(l_{t+\mathrm{d}t})$ being taken from A_1 to B_1 and A_2 to B_2 respectively, whilst (n) is the normal direction to (δS) such that the vectors d$\mathbf{s}$, $\mathbf{v}$ and (n) on (l_t) have the same orientation as the axes. The integrals over the small curves (A_2B_2) and (B_1A_1) can be replaced by single elements, i.e. the products of integrand and length of path; we have for these the scalar products of $\mathbf{A}$ and the small displacements $\mathbf{v}$ dt:

$$\mathbf{A}^{(2)}\cdot\mathbf{v}^{(2)}\,\mathrm{d}t \quad\text{and}\quad -\mathbf{A}^{(1)}\cdot\mathbf{v}^{(1)}\,\mathrm{d}t,$$

where the $(-)$ sign is used because integration over curve B_1A_1 is carried out from B_1 to A_1, i.e. in the opposite direction to $\mathbf{v}$, whilst the superscripts denote values corresponding to the ends A_1 and A_2.

An elementary area dS is given by

$$\mathrm{d}S = |\,\mathrm{d}\mathbf{s}\times\mathbf{v}\,|\,\mathrm{d}t,$$

and the normal (n) to the surface has the same direction as the vector $\mathbf{ds}\times\mathbf{v}$, so that clearly:

$$\operatorname{curl}_n \mathbf{A}\, \mathrm{d}S = (\mathbf{ds}\times\mathbf{v})\cdot \operatorname{curl}\mathbf{A}\, \mathrm{d}t = (\mathbf{v}\times\operatorname{curl}\mathbf{A})\cdot \mathbf{ds}\, \mathrm{d}t,$$

and (122) gives:

$$\int\limits_{(l_{t+\mathrm{d}t})} A_s\, \mathrm{d}s - \int\limits_{(l_t)} A_s\, \mathrm{d}s = \mathbf{A}^{(2)}\cdot\mathbf{v}^{(2)}\,\mathrm{d}t - \mathbf{A}^{(1)}\cdot\mathbf{v}^{(1)}\,\mathrm{d}t + \mathrm{d}t\int\limits_{(l_t)}(\operatorname{curl}\mathbf{A}\times\mathbf{v})_s\,\mathrm{d}s.$$

On dividing both sides by $\mathrm{d}t$, passing to the limit and adding term (121), we get the required derivative, where we write simply (l) instead of (l_t):

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{(l)} A_s\,\mathrm{d}s = \mathbf{A}^{(2)}\cdot\mathbf{v}^{(2)} - \mathbf{A}^{(1)}\cdot\mathbf{v}^{(1)} + \int\limits_{(l)}\left[\frac{\partial A_s}{\partial t} + (\operatorname{curl}\mathbf{A}\times\mathbf{v})_s\right]\mathrm{d}s. \tag{123}$$

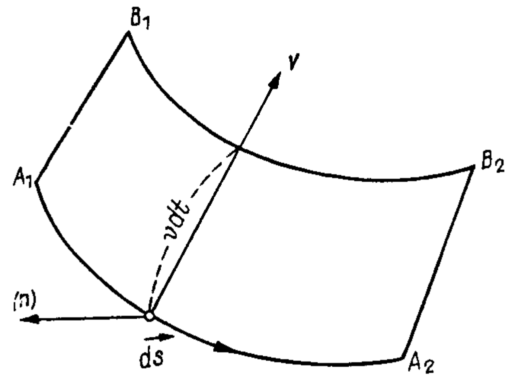

FIG. 99

If (l) is a closed curve, the terms outside the integral fall out, and we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{(l)} A_s\,\mathrm{d}s = \int\limits_{(l)}\left[\frac{\partial A_s}{\partial t} + (\operatorname{curl}\mathbf{A}\times\mathbf{v})_s\right]\mathrm{d}s. \tag{124}$$

This expression can be obtained more simply by transforming the line integral by means of Stokes' formula, then using (120).

We further consider the circulation of the velocity along the moving contour (l). By (123):

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{(l)} v_s\,\mathrm{d}s = \mathbf{v}^{(2)}\cdot\mathbf{v}^{(2)} - \mathbf{v}^{(1)}\cdot\mathbf{v}^{(1)} + \int\limits_{(l)}\left[\frac{\partial v_s}{\partial t} + (\operatorname{curl}\mathbf{v}\times\mathbf{v})_s\right]\mathrm{d}s =$$

$$= |\mathbf{v}^{(2)}|^2 - |\mathbf{v}^{(1)}|^2 + \int\limits_{(l)}\left[\frac{\partial v_s}{\partial t} + (\operatorname{curl}\mathbf{v}\times\mathbf{v})_s\right]\mathrm{d}s. \tag{125}$$

The component of $\operatorname{curl}\mathbf{v}\times\mathbf{v}$ along OX will be:

$$(\operatorname{curl}\mathbf{v}\times\mathbf{v})_x = \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right)v_z - \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right)v_y.$$

On removing the brackets and adding and subtracting $v_x \partial v_x/\partial x$, we can write:

$$(\operatorname{curl} \mathbf{v} \times \mathbf{v})_x = \frac{\partial v_x}{\partial x} v_x + \frac{\partial v_x}{\partial y} v_y + \frac{\partial v_x}{\partial z} v_z - \\ - \left(\frac{\partial v_x}{\partial x} v_x + \frac{\partial v_y}{\partial x} v_y + \frac{\partial v_z}{\partial x} v_z \right),$$

and hence we easily obtain, using (115):

$$\operatorname{curl} \mathbf{v} \times \mathbf{v} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} - \frac{\partial \mathbf{v}}{\partial t} - \frac{1}{2} \operatorname{grad} |\mathbf{v}|^2 = \mathbf{w} - \frac{\partial \mathbf{v}}{\partial t} - \frac{1}{2} \operatorname{grad} |\mathbf{v}|^2,$$

where $\mathbf{w}$ is the acceleration vector. We have on substituting in (125):

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{(l)} v_s \, \mathrm{d}s = |\mathbf{v}^{(2)}|^2 - |\mathbf{v}^{(1)}|^2 + \int_{(l)} (w_s - \frac{1}{2} \operatorname{grad}_s |\mathbf{v}|^2) \, \mathrm{d}s = \\ = \frac{1}{2} [|\mathbf{v}^{(2)}|^2 - |\mathbf{v}^{(1)}|^2] + \int_{(l)} w_s \, \mathrm{d}s, \qquad (126)$$

since clearly:

$$\int_{(l)} \operatorname{grad}_s |\mathbf{v}|^2 \, \mathrm{d}s = |\mathbf{v}^{(2)}|^2 - |\mathbf{v}^{(1)}|^2.$$

CHAPTER V

FOUNDATIONS OF DIFFERENTIAL GEOMETRY

§ 12. Curves on a plane and in space

121. The curvature of a plane curve; the evolute. The present chapter deals with the basic theory of curves and surfaces; it starts with plane curves, after which we pass to curves in space and surfaces. Our account makes use of vectors, so that the reader must carefully bear in mind the initial sections of the previous chapter up to and including [107], which deals with the differentiation of vectors. We start by proving a lemma:

LEMMA. *If* $\mathbf{A}$ *is a vector of unit length (a unit vector) which depends on a scalar parameter* t, *we have* $\mathrm{d}\mathbf{A}/\mathrm{d}t \cdot \mathbf{A} = 0$, i.e. $\mathrm{d}\mathbf{A}/\mathrm{d}t \perp \mathbf{A}$. In fact, $\mathbf{A} \cdot \mathbf{A} = 1$ by hypothesis, and we obtain on differentiating this equation with respect to t:

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} \cdot \mathbf{A} + \mathbf{A} \cdot \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = 0,$$

or, by the independence of a scalar product on the order of its factors:

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} \cdot \mathbf{A} = 0, \quad \text{i.e.} \quad \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} \perp \mathbf{A},$$

where the orthogonality condition obviously only has a meaning in the case when $\mathrm{d}\mathbf{A}/\mathrm{d}t$ differs from zero.

Here and in future, we always assume the existence and continuity of the derivatives mentioned in the text.

Let (L) be a plane curve and let the position of a variable point M on the curve be described by a scalar parameter t. The curve may be characterized by the radius vector $\mathbf{r}(t)$ from some fixed point O to a variable point of the curve (Fig. 100). The derivative $\mathrm{d}\mathbf{r}/\mathrm{d}t$ gives a vector directed along the tangent to the curve, as we saw in [107], whilst if the length of arc s, measured from some definite point of the

curve in a definite direction, is taken as the parameter, $d\mathbf{r}/ds$ is the *unit tangential vector* $\mathbf{t}$, the direction of which coincides with the direction of increase of parameter s along the curve:

$$\frac{d\mathbf{r}}{ds} = \mathbf{t}. \tag{1}$$

The derivative of the unit tangential vector with respect to s is called the *curvature vector:*

$$\mathbf{N} = \frac{d\mathbf{t}}{ds}. \tag{2}$$

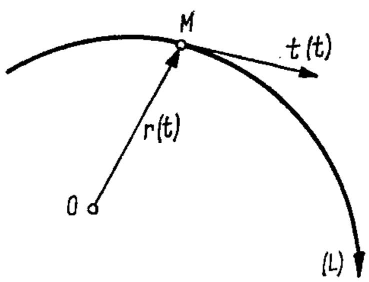

FIG. 100

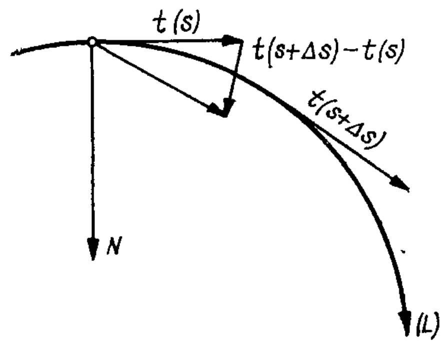

FIG. 101

The length of this vector characterizes the rapidity of change in the direction of the vector $\mathbf{t}$, and is called the *curvature*.

The curvature vector is perpendicular to the tangent by the lemma just proved, i.e. is directed along the normal.

Furthermore, it follows at once from the definition that it is directed towards the concave side of the curve, since the difference $\mathbf{t}(s + \Delta s) - \mathbf{t}(s)$, with $\Delta s > 0$, is directed towards this side (Fig. 101).

The length of vector $\mathbf{N}$ is called the curvature, as already mentioned, and if we introduce the notation

$$|\mathbf{N}| = \frac{1}{\varrho}, \tag{3}$$

the inverse of the curvature, ϱ, is called the *radius of curvature*. We now write $\mathbf{n}$ as the unit curvature vector, i.e. the vector of unit length with the same direction as $\mathbf{N}$.

If the length $|\mathbf{N}| = 0$, we must take $\varrho = \infty$, whilst $\mathbf{n}$ is not defined. For instance, if (L) is a straight line, $|\mathbf{N}| = 0$ at every point of it, and we can choose either of the two directions for the normal lying in the plane concerned. We shall assume in future that $|\mathbf{N}| \neq 0$.

We have by (3):

$$\mathbf{N} = \frac{1}{\varrho}\mathbf{n}. \tag{4}$$

Let us mark off along the direction of $\mathbf{n}$, i.e. along the normal direction towards the concave side, the segment MC, equal to the radius of curvature at the point M of the curve (Fig. 102). Its end C is called the centre of curvature of the curve at the point M. If M moves along the curve (L), C varies and describes a curve (L_1), called the *evolute of curve* (L), i.e. *the evolute of a curve is defined as the locus of its centre of curvature.*

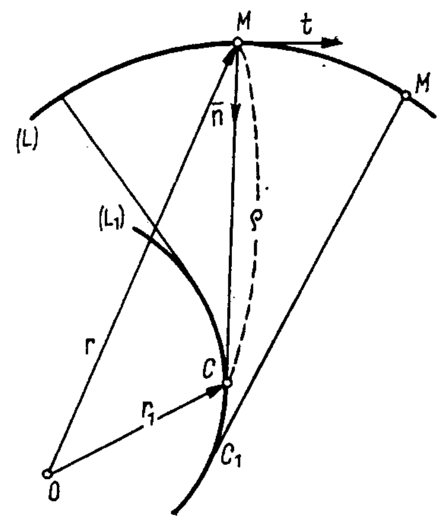

FIG. 102

We must find the derivative $\mathrm{d}\mathbf{n}/\mathrm{d}s$ for the sake of what follows. Since $\mathbf{n}$ is a unit vector, $\mathrm{d}\mathbf{n}/\mathrm{d}s \perp \mathbf{n}$ i.e. $\mathrm{d}\mathbf{n}/\mathrm{d}s$ is directed along the tangent. We obtain on differentiating the obvious equation $\mathbf{t}\cdot\mathbf{n} = 0$ with respect to s:

$$\mathbf{N}\cdot\mathbf{n} + \mathbf{t}\cdot\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s} = 0.$$

But $\mathbf{N}$ and $\mathbf{n}$ have the same direction, and $\mathbf{N}\cdot\mathbf{n} = 1/\varrho$ by (4), so that it follows from the above equation that $\mathbf{t}\cdot \mathrm{d}\mathbf{n}/\mathrm{d}s = -1/\varrho$. We see from the fact that $\mathbf{t}$ and $\mathrm{d}\mathbf{n}/\mathrm{d}s$ are parallel that $\mathrm{d}\mathbf{n}/\mathrm{d}s$ is in the opposite direction to $\mathbf{t}$ and has a length $1/\varrho$, i.e.

$$\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s} = -\frac{1}{\varrho}\mathbf{t}. \tag{5}$$

Let $\mathbf{r}$ be the radius vector and s the length of arc of the curve (L) as above, and let $\mathbf{r}_1$ and s_1 be the corresponding magnitudes for the evolute (L_1). On differentiating with respect to s the equation (Fig. 102)

$$\mathbf{r}_1 = \mathbf{r} + \varrho\mathbf{n}$$

we get

$$\frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}s} = \mathbf{t} + \frac{\mathrm{d}\varrho}{\mathrm{d}s}\mathbf{n} + \varrho\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s},$$

or, by (5):

$$\frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}s} = \mathbf{t} + \frac{\mathrm{d}\varrho}{\mathrm{d}s}\mathbf{n} - \mathbf{t}, \quad \text{i.e.} \quad \frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}s} = \frac{\mathrm{d}\varrho}{\mathrm{d}s}\mathbf{n}. \tag{6}$$

The right-hand side of this equation is a vector directed along the normal to (L), whilst the left-hand side is a vector directed along the tangent to the evolute; hence the normal to (L) is parallel to the tangent to the evolute. But both these latter past through the same point C and are thus identical, leading us to the first property of evolutes: *a normal to a curve is tangential to the evolute at its corresponding point.*

On recalling the definition of envelope, we can state a second property of evolutes: *the evolute is the envelope of normals to the curve.*

It seems natural to take the length of arc s_1 as the parameter for the evolute, and by the rule for differentiation of a function of a function:

$$\frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}s} = \frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}s_1} \cdot \frac{\mathrm{d}s_1}{\mathrm{d}s} = \frac{\mathrm{d}s_1}{\mathrm{d}s}\,\mathbf{t}_1,$$

where $\mathbf{t}_1$ is the unit tangential vector to the evolute. We get on substituting in (6):

$$\frac{\mathrm{d}s_1}{\mathrm{d}s}\,\mathbf{t}_1 = \frac{\mathrm{d}\varrho}{\mathrm{d}s}\,\mathbf{n},$$

and we see by comparing the lengths of the vectors on the two sides of the equation that

$$\left|\frac{\mathrm{d}s_1}{\mathrm{d}s}\right| = \left|\frac{\mathrm{d}\varrho}{\mathrm{d}s}\right|, \quad \text{i.e.} \quad |\mathrm{d}s_1| = |\mathrm{d}\varrho|.$$

If we assume for simplicity that ϱ and s_1 are increasing over the. sections of the curve and evolute concerned, we can write $\mathrm{d}s_1 = \mathrm{d}\varrho$.† Integration of this relationship over the section concerned shows that the increment in length of arc of the evolute is the same as the increment in the radius of curvature of the initial curve. Hence we have a third property of evolutes: *the increment in the radius of curvature of the curve over a section of monotonic change is equal to the length of arc between the corresponding points of the evolute.* In the case of Fig. 102, this property is expressed by: $M_1C_1 - MC = \smile CC_1$.

We take definite axes OX, OY on the plane and let φ be the angle that the tangential direction $\mathbf{t}$ forms with OX. We express the unit vector in terms of its components:

$$\mathbf{t} = \cos\varphi\mathbf{i} + \sin\varphi\mathbf{j},$$

† On the assumption that ϱ is varying monotonically, we can always choose the direction for measuring s_1 such that s_1 is increasing monotonically with ϱ.

where **i** and **j** are the unit vectors along OX, OY. We obtain on differentiating with respect to s:

$$\mathbf{N} = -\sin\varphi \frac{d\varphi}{ds}\,\mathbf{i} + \cos\varphi \frac{d\varphi}{ds}\,\mathbf{j},$$

whence the square of the length of the curvature vector is given by:

$$\frac{1}{\varrho^2} = \left(-\sin\varphi \frac{d\varphi}{ds}\right)^2 + \left(\cos\varphi \frac{d\varphi}{ds}\right)^2 \quad \text{or} \quad \frac{1}{\varrho} = \left|\frac{d\varphi}{ds}\right|.$$

This is the same expression for the curvature as obtained in [I, 71].

Let the equation of (L) be given explicitly as

$$y = f(x). \tag{7}$$

The equation of the family of normals to the curve will be

$$Y - y = -\frac{1}{y'}(X - x) \quad \text{or} \quad (X - x) + y'(Y - y) = 0. \tag{8}$$

Here, (X, Y) are the current coordinates of the normal, (x, y) are the coordinates of the point M of (L), and y is the function (7) of x. The role of parameter in equation (8) of the family of normals is played by the abscissa x of a variable point of the curve. On applying to (8) the usual rule for finding the envelope [10], two equations have to be written: (8), and a new equation obtained by differentiating (8) with respect to the parameter x:

$$\left.\begin{aligned} (X - x) + y'(Y - y) &= 0; \\ -1 + y''(Y - y) - y'^2 &= 0. \end{aligned}\right\} \tag{9}$$

Elimination of x from these equations gives an equation connecting X and Y. This is in fact the equation of the normal envelope, i.e. the equation of the evolute. An alternative procedure is to solve system (9) for X and Y in terms of the parameter x, so as to obtain the parametric equations of the evolute:

$$X = x - \frac{y'(1 + y'^2)}{y''}; \quad Y = y + \frac{1 + y'^2}{y''}. \tag{10}$$

If the equation of (L) is itself given parametrically, the derivatives of y with respect to x in (10) have to be expressed in terms of the differentials of the variables [I, 74]:

$$y' = \frac{dy}{dx}; \quad y'' = \frac{d\left(\frac{dy}{dx}\right)}{dx} = \frac{d^2y\,dx - d^2x\,dy}{dx^3}.$$

Substitution of these expressions in (10) gives us the parametric equations of the evolute in this case as:

$$X = x - \frac{dy(dx^2 + dy^2)}{d^2y\,dx - d^2x\,dy}\ ;\ \ Y = y + \frac{dx(dx^2 + dy^2)}{d^2y\,dx - d^2x\,dy}\,. \tag{11}$$

Examples. **1.** We find the evolute of the ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad\qquad (a > b).$$

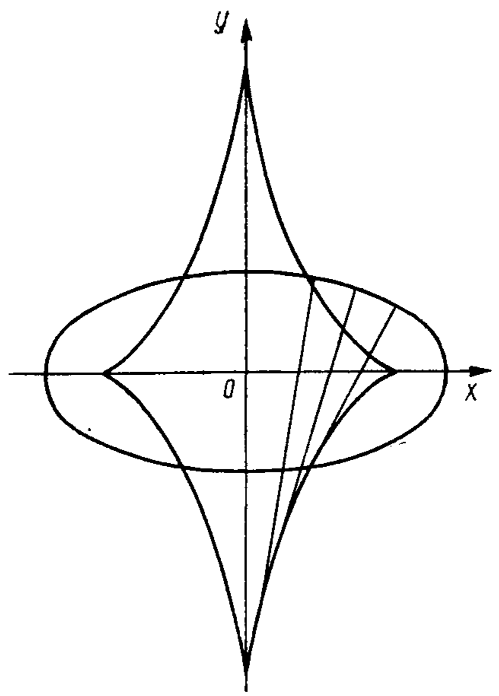

FIG. 103

On writing the equation of the ellipse in the parametric form

$$x = a\cos t;\quad y = b\sin t$$

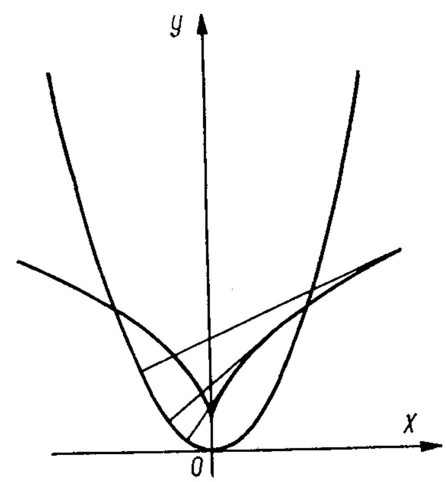

FIG. 104

and substituting in (11), we find after simplification:

$$X = \frac{a^2 - b^2}{a}\cos^3 t,$$

$$Y = -\frac{a^2 - b^2}{b}\sin^3 t.$$

We eliminate parameter t from these two equations. We multiply the first by a and the second by b, raise to the power 2/3 and add; this gives us the evolute of the ellipse in the implicit form:

$$a^{\frac{2}{3}}X^{\frac{2}{3}} + b^{\frac{2}{3}}Y^{\frac{2}{3}} = (a^2 - b^2)^{\frac{2}{3}}.$$

The evolute is easily plotted by using these equations. The radius of curvature attains its least and greatest values at the vertices of the ellipse and the corresponding points of the evolute are singular points (cusps) (Fig. 103).

2. We find the evolute of the parabola $y = ax^2$. We easily find, on using equations (10):

$$X = -4a^2x^3, \quad Y = \frac{1}{2a} + 3ax^2.$$

Elimination of the parameter x now gives us the evolute of the parabola as explicitly (Fig. 104):

$$Y = \frac{1}{2a} + \frac{3}{2\sqrt{2a}} X^{\frac{2}{3}}.$$

3. We take the cycloid:

$$x = a(t - \sin t); \quad y = a(1 - \cos t).$$

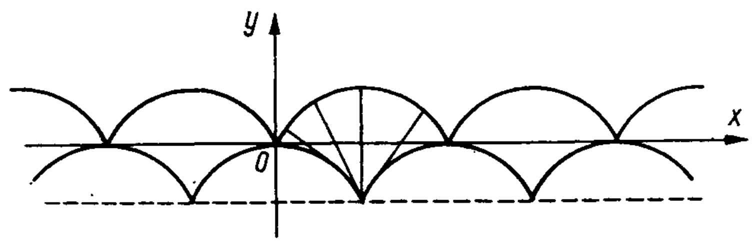

FIG. 105

We use equations (11) to find the parametric equations of the evolute as

$$X = a(t + \sin t); \quad Y = -a(1 - \cos t).$$

This may easily be shown to be the same curve as the original cycloid, except for being displaced relative to the axes (Fig. 105). We simply put $t = \tau - \pi$, and write the last equations as:

$$X + a\pi = a(\tau - \sin \tau),$$

$$Y + 2a = a(1 - \cos \tau),$$

whence our statement follows immediately.

122. Involutes. The curve (L) itself is called the *involute* of its evolute (L_1). A rule for drawing the involute of a given evolute is easily derived from the properties of evolutes. If C is a variable point of (L_1) and s_1 is its length of arc, we cut off a section $\overline{CM} = s_1 + a$, where a is a constant, in the negative direction along the tangent to (L_1) at C, and consider the locus (L) of the end M. This locus may easily be shown to be the required involute (Fig. 106). It is sufficient

to show that CM is normal to the curve (L). Let $\mathbf{r}$ and $\mathbf{r}_1$ be the radius vectors to (L) and (L_1), as above, and let $\mathbf{t}_1$ be the unit tangential vector to (L_1). By construction:

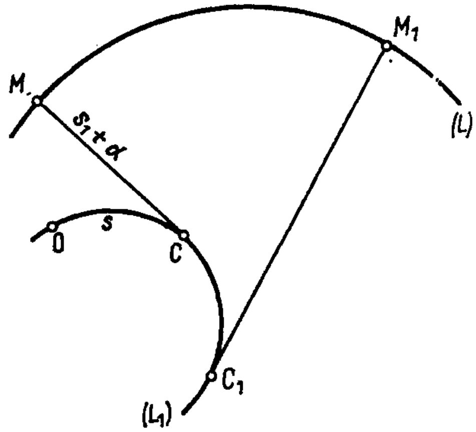

FIG. 106

$$\mathbf{r} = \mathbf{r}_1 - (s_1 + a)\,\mathbf{t}_1,$$

whence differentiation with respect to s_1 gives:

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s_1} = \mathbf{t}_1 - \mathbf{t}_1 - (s_1 + a)\frac{\mathrm{d}t_1}{\mathrm{d}s_1},$$

$$\text{i.e. } \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s_1} = -(s_1 + a)\frac{\mathrm{d}\mathbf{t}_1}{\mathrm{d}s_1}.$$

It is clear from this that the vector $\mathrm{d}\mathbf{r}/\mathrm{d}s_1$, parallel to the tangent to (L), is at the same time parallel to the vector $\mathrm{d}\mathbf{t}_1/\mathrm{d}s_1$, i.e. to the normal to (L_1); hence it follows that the tangent $\overline{CM}$ to (L_1) is normal to (L).

We can assign arbitrary values to the constant a in the expression $\overline{CM} = s_1 + a$, so that an infinite number of involutes can be obtained for a given evolute. It follows from the method of construction that any two involutes have a common normal, the intercept of which between the curves maintains a constant length equal to the difference between the values of the constant a taken. Two such curves are said to be *parallel*.

123. The natural equation† of a curve. The curvature along any curve is a definite function of the length of arc:

$$\frac{1}{\varrho} = f(s). \tag{12}$$

We show that, conversely, a single curve is defined by any equation of the type (12). We fix the direction of the x axis arbitrarily and let φ be the angle between the tangent to the curve and this axis. We know that $1/\varrho = \pm\,\mathrm{d}\varphi/\mathrm{d}s$, and (12) gives:

$$\frac{\mathrm{d}\varphi}{\mathrm{d}s} = \pm f(s),$$

whence

$$\varphi = \pm \int_0^s f(s)\,\mathrm{d}s + C.$$

† The natural equation is also known as the intrinsic equation.

We may assume that the direction of the x axis coincides with the direction of the tangent at $s = 0$, so that we may take $C = 0$ in the last equation, i.e. we obtain as an expression for the angle φ:

$$\varphi = \pm F(s), \quad \text{where} \quad F(s) = \int_0^s f(s)\,\mathrm{d}s.$$

We know moreover [I, 70] that

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \cos\varphi; \qquad \frac{\mathrm{d}y}{\mathrm{d}s} = \sin\varphi,$$

whence, by the previous equation:

$$x = \int_0^s \cos[F(s)]\,\mathrm{d}s + C_1,$$

$$y = \pm \int_0^s \sin[F(s)]\,\mathrm{d}s + C_2.$$

If we locate the origin at the point of the curve for which $s = 0$, we must take $C_1 = C_2 = 0$, and we get the fully defined curve

$$x = \int_0^s \cos[F(s)]\;\mathrm{d}s; \quad y = \pm \int_0^s \sin[F(s)]\,\mathrm{d}s. \tag{12$_1$}$$

The $\pm$ sign merely gives symmetry with respect to the x axis.

We have now shown that (12) can correspond to a curve defined in the above sense and that equations (12_1) must specify the curve parametrically with the coordinate system chosen. It may easily be verified that the curvature of the curve defined by (12_1) in fact has the value given by (12).

We speak of (12) as the *natural equation of the curve*, in the sense that it is independent of the particular coordinate system chosen, whilst a fully defined curve corresponds to it (except for symmetry).

Examples. **1.** If (12) has the form $1/\varrho = C$, i.e. the radius of curvature is constant, we know that a circle satisfies the equation [I, 71]. It follows from the above that *a circle is the only curve with constant radius of curvature.*

2. Let the curvature $1/\varrho$ be proportional to the length of arc:

$$\frac{1}{\varrho} = 2as,$$

where $2a$ is the positive coefficient of proportionality. The above working gives in this case:

$$x = \int_0^s \cos(as^2)\,\mathrm{d}s; \qquad y = \int_0^s \sin(as^2)\,\mathrm{d}s. \tag{13}$$

We can say from the convergence of the integrals [83]:

$$\int_0^\infty \cos(as^2)\,\mathrm{d}s; \qquad \int_0^\infty \sin(as^2)\,\mathrm{d}s$$

that on indefinite increase of s the curve will tend to a point of the plane with coordinates equal to the values of these integrals, the point being approached by spiralling about it (Fig. 107). If s is assigned negative values in expressions (13), we get the portion of the curve contained in the third quadrant of the axes. The curve obtained here is known as a Cornu spiral and is encountered in optics.

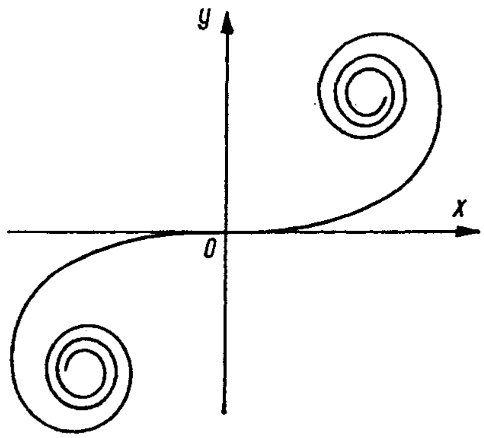

FIG. 107

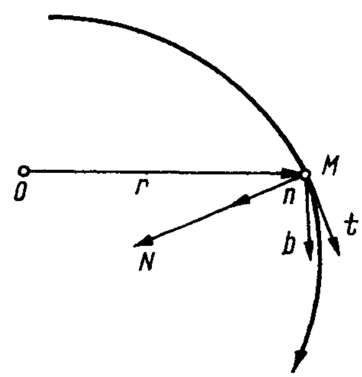

FIG. 108

124. The fundamental elements of curves in space. A curve (L) in space can be defined by specifying the radius vector $\mathbf{r}(t)$ from the origin to a variable point M of the curve (Fig. 108). If we take the length of arc s as the parameter t and differentiate $\mathbf{r}$ with respect to s, we get the unit tangential vector to the curve [107]:

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \mathbf{t}. \tag{14}$$

The derivative of $\mathbf{t}$ with respect to s is known as the *curvature vector:*

$$\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \mathbf{N}, \tag{15}$$

and the length of this defines the *curvature* $1/\varrho$, the inverse ϱ being called the *radius of curvature*. The vector $\mathbf{N}$ is perpendicular to $\mathbf{t}$, as in the case of a plane curve, and the direction of $\mathbf{N}$ is known as the direction of the *principal normal to the curve*. If $\mathbf{n}$ denotes the unit vector in the direction of the principal normal, we can write:

$$\mathbf{N} = \frac{1}{\varrho}\mathbf{n}. \tag{16}$$

Now let $\mathbf{b}$ be the unit vector perpendicular to $\mathbf{t}$ and $\mathbf{n}$:

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}. \tag{17}$$

This is known as the *unit binormal vector*.

The three unit vectors $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{b}$, having the same orientation as the coordinate axes, are said to form a *variable triad related to the curve* (L). If (L) is a plane curve, vectors $\mathbf{t}$ and $\mathbf{n}$ lie in the plane and the unit binormal vector $\mathbf{b}$ is therefore a constant vector of unit length perpendicular to the plane. In the case of a non-plane curve, the derivative $\mathrm{d}\mathbf{b}/\mathrm{d}s$ characterizes the deviation of the curve from the plane form and is called the *torsional vector*. We show that *the torsional vector is parallel to the principal normal*. From (17):

$$\frac{\mathrm{d}\mathbf{b}}{\mathrm{d}s} = \mathbf{N} \times \mathbf{n} + \mathbf{t} \times \frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s}.$$

But $\mathbf{N}$ and $\mathbf{n}$ have the same direction so that their vector product is zero, i.e.

$$\frac{\mathrm{d}\mathbf{b}}{\mathrm{d}s} = \mathbf{t} \times \frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s}, \tag{18}$$

whence it follows that $\mathrm{d}\mathbf{b}/\mathrm{d}s$ and $\mathbf{t}$ are perpendicular. We know on the other hand that $\mathrm{d}\mathbf{b}/\mathrm{d}s$, the derivative of a unit vector, must be perpendicular to $\mathbf{b}$ itself. Hence $\mathrm{d}\mathbf{b}/\mathrm{d}s$, erpendicular to $\mathbf{t}$ and $\mathbf{b}$, must in fact be parallel to $\mathbf{n}$, and we can write

$$\frac{\mathrm{d}\mathbf{b}}{\mathrm{d}s} = \frac{1}{\tau}\mathbf{n}, \tag{19}$$

where the numerical coefficient $1/\tau$ is called the *torsion of the curve*, the inverse τ being the *radius of torsion* or *radius of second curvature*. Unlike the curvature $1/\varrho$, $1/\tau$ can be either positive or negative. Needless to say, the existence of the tangential, curvature and torsional vectors is bound up with the existence of the derivatives in terms of which they are expressed.

We now obtain formulae for calculating the curvature and torsion. On taking Cartesian axes OX, OY, OZ, with the corresponding unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, we can write:

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}; \quad \mathbf{t} = \frac{\mathrm{d}x}{\mathrm{d}s}\mathbf{i} + \frac{\mathrm{d}y}{\mathrm{d}s}\mathbf{j} + \frac{\mathrm{d}z}{\mathrm{d}s}\mathbf{k};$$

$$\mathbf{N} = \frac{\mathrm{d}^2 x}{\mathrm{d}s^2}\mathbf{i} + \frac{\mathrm{d}^2 y}{\mathrm{d}s^2}\mathbf{j} + \frac{\mathrm{d}^2 z}{\mathrm{d}s^2}\mathbf{k},$$

whence we obtain for the length of the vector **N**:

$$\frac{1}{\varrho^2} = \left(\frac{d^2x}{ds^2}\right)^2 + \left(\frac{d^2y}{ds^2}\right)^2 + \left(\frac{d^2z}{ds^2}\right)^2. \tag{20}$$

It follows from (19) that the torsion $1/\tau$ can be expressed as a scalar product:

$$\frac{1}{\tau} = \frac{d\mathbf{b}}{ds} \cdot \mathbf{n}$$

or, by (18),

$$\frac{1}{\tau} = \left(\mathbf{t} \times \frac{d\mathbf{n}}{ds}\right) \cdot \mathbf{n}.$$

On substituting for **n** from (16):

$$\mathbf{n} = \varrho\mathbf{N},$$

we get:

$$\frac{1}{\tau} = \left(\mathbf{t} \times \frac{d(\varrho\mathbf{N})}{ds}\right) \cdot \varrho\mathbf{N} = \left[\mathbf{t} \times \left(\frac{d\varrho}{ds}\mathbf{N} + \varrho\frac{d\mathbf{N}}{ds}\right)\right] \cdot \varrho\mathbf{N} =$$

$$= \varrho\frac{d\varrho}{ds}(\mathbf{t} \times \mathbf{N}) \cdot \mathbf{N} + \varrho^2 \cdot \left(\mathbf{t} \times \frac{d\mathbf{N}}{ds}\right) \cdot \mathbf{N}.$$

The vector product $\mathbf{t} \times \mathbf{N}$ is perpendicular to **N**, so that the first term in the last expression is zero; hence

$$\frac{1}{\tau} = \varrho^2 \left(\mathbf{t} \times \frac{d\mathbf{N}}{ds}\right) \cdot \mathbf{N},$$

or, on transposing the factors in the vector product:

$$\frac{1}{\tau} = -\varrho^2 \left(\frac{d\mathbf{N}}{ds} \times \mathbf{t}\right) \cdot \mathbf{N}.$$

We get finally, on carrying out a cyclic change of the vectors and using (14) and (15):

$$\frac{1}{\tau} = -\varrho^2 \left(\frac{d\mathbf{r}}{ds} \times \frac{d^2\mathbf{r}}{ds^2}\right) \cdot \frac{d^3\mathbf{r}}{ds^3}. \tag{21}$$

It may be noted that the coefficient of $(-\varrho^2)$ is the volume of the parallelepiped formed by the vectors $d\mathbf{r}/ds$, $d^2\mathbf{r}/ds^2$, $d^3\mathbf{r}/ds^3$ [105].

We return to expression (20) for the curvature. It is assumed in this that the coordinates x, y, z are given as functions of the length of arc. We now obtain a new form of (20), suitable for a curve given in any parametric form. We shall need to express the derivative of the coordinate with respect to the length of arc in terms of the differential of the coordinate. Differentiation of the expression

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{22}$$

gives us:

$$\mathrm{d}s\,\mathrm{d}^2s = \mathrm{d}x\,\mathrm{d}^2x + \mathrm{d}y\,\mathrm{d}^2y + \mathrm{d}z\,\mathrm{d}^2z. \tag{23}$$

We have, moreover [I, 74]:

$$\frac{\mathrm{d}^2x}{\mathrm{d}s^2} = \frac{\mathrm{d}^2x\,\mathrm{d}s - \mathrm{d}^2s\,\mathrm{d}x}{\mathrm{d}s^3}; \quad \frac{\mathrm{d}^2y}{\mathrm{d}s^2} = \frac{\mathrm{d}^2y\,\mathrm{d}s - \mathrm{d}^2s\,\mathrm{d}y}{\mathrm{d}s^3};$$

$$\frac{\mathrm{d}^2z}{\mathrm{d}s^2} = \frac{\mathrm{d}^2z\,\mathrm{d}s - \mathrm{d}^2s\,\mathrm{d}z}{\mathrm{d}s^3}. \tag{24}$$

We obtain on making these substitutions in (20):

$$\frac{1}{\varrho^2} =$$

$$= \frac{\mathrm{d}s^2[(\mathrm{d}^2x)^2+(\mathrm{d}^2y)^2+(\mathrm{d}^2z)^2]-2\mathrm{d}s\mathrm{d}^2s(\mathrm{d}x\mathrm{d}^2x+\mathrm{d}y\mathrm{d}^2y+\mathrm{d}z\mathrm{d}^2z)+(\mathrm{d}^2s)^2(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2)}{\mathrm{d}s^6}$$

or, by (22) and (23):

$$\frac{1}{\varrho^2} =$$

$$= \frac{(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2)[(\mathrm{d}^2x)^2+\mathrm{d}^2y)^2+(\mathrm{d}^2z)^2]-(\mathrm{d}x\,\mathrm{d}^2x+\mathrm{d}y\,\mathrm{d}^2y+\mathrm{d}z\,\mathrm{d}^2z)^2}{\mathrm{d}s^6}. \tag{25}$$

We now recall from [104] an elementary algebraic identity that we shall need:

$$(a^2+b^2+c^2)(a_1^2+b_1^2+c_1^2)-(aa_1+bb_1+cc_1)^2 =$$

$$= (bc_1-cb_1)^2+(ca_1-ac_1)^2+(ab_1-ba_1)^2. \tag{26}$$

On applying this identity to the numerator of the right-hand side of (25), we can write the square of the curvature in the final form:

$$\frac{1}{\varrho^2} = \frac{A^2+B^2+C^3}{(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2)^3}, \tag{27}$$

where

$$A = \mathrm{d}y\,\mathrm{d}^2z - \mathrm{d}z\,\mathrm{d}^2y; \qquad B = \mathrm{d}z\,\mathrm{d}^2x - \mathrm{d}x\,\mathrm{d}^2z;$$

$$C = \mathrm{d}x\,\mathrm{d}^2y - \mathrm{d}y\,\mathrm{d}^2x.$$

If (L) is the trajectory of a moving point, the velocity vector will be defined by

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\mathrm{d}s}{\mathrm{d}t}\mathbf{t}.$$

Further differentiation with respect to time gives us the acceleration vector

$$\mathbf{w} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\mathrm{d}^2s}{\mathrm{d}t^2}\mathbf{t} + \frac{\mathrm{d}s}{\mathrm{d}t}\cdot\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}t},$$

or by (15) and (16):

$$\mathbf{w} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\,\mathbf{t} + \frac{\mathrm{d}s}{\mathrm{d}t}\cdot\frac{\mathrm{d}s}{\mathrm{d}t}\cdot\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\,\mathbf{t} + \frac{v^2}{\varrho}\,n \quad \left(v = \frac{\mathrm{d}s}{\mathrm{d}t}\right),$$

whence it is clear that the acceleration vector has a component along the tangent equal to $\mathrm{d}^2s/\mathrm{d}t^2$ and one along the normal equal to v^2/ϱ, whilst the component along the binormal is zero.

125. Frenet's formulae. We introduce a notation for the direction-cosines of the axes of a movable triad with respect to fixed coordinate axes, as indicated in the accompanying table.

	X	Y	Z
t	α	β	γ
n	α_1	β_1	γ_1
b	α_2	β_2	γ_2

Frenet's formulae give the derivatives of the nine direction-cosines written with respect to s.

The components of unit vector $\mathbf{t}$ are α, β, γ, and the expression

$$\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \mathbf{N} = \frac{1}{\varrho}\,\mathbf{n}$$

leads to the first three of Frenet's formulae:

$$\frac{\mathrm{d}\alpha}{\mathrm{d}s} = \frac{\alpha_1}{\varrho}\,; \quad \frac{\mathrm{d}\beta}{\mathrm{d}s} = \frac{\beta_1}{\varrho}\,; \quad \frac{\mathrm{d}\gamma}{\mathrm{d}s} = \frac{\gamma_1}{\varrho}\,. \tag{28}$$

Similarly, (19) leads to the three further formulae:

$$\frac{\mathrm{d}\alpha_2}{\mathrm{d}s} = \frac{\alpha_1}{\tau}\,; \quad \frac{\mathrm{d}\beta_2}{\mathrm{d}s} = \frac{\beta_1}{\tau}\,; \quad \frac{\mathrm{d}\gamma_2}{\mathrm{d}s} = \frac{\gamma_1}{\tau}\,. \tag{28$_1$}$$

Consideration of the movable triad gives us directly $\mathbf{n} = -\,\mathbf{t}\times\mathbf{b}$, and we obtain by differentiation with respect to s:

$$\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s} = -\frac{1}{\varrho}\,\mathbf{n}\times\mathbf{b} - \frac{1}{\tau}\,\mathbf{t}\times\mathbf{n} = -\frac{1}{\varrho}\,\mathbf{t} - \frac{1}{\tau}\,\mathbf{b}.$$

This gives the final three formulae:

$$\frac{\mathrm{d}\alpha_1}{\mathrm{d}s} = -\frac{\alpha}{\varrho} - \frac{\alpha_2}{\tau}\,; \quad \frac{\mathrm{d}\beta_1}{\mathrm{d}s} = -\frac{\beta}{\varrho} - \frac{\beta_2}{\tau}\,;$$

$$\frac{\mathrm{d}\gamma_1}{\mathrm{d}s} = -\frac{\gamma}{\varrho} - \frac{\gamma_2}{\tau}\,. \tag{28$_2$}$$

We can easily show by using (28) that *if the curvature* $1/\varrho$ *is zero along a curve* (L), (L) *is a straight line*. The identity $1/\varrho = 0$ gives, in fact,

$$\frac{\mathrm{d}\alpha}{\mathrm{d}s} = \frac{\mathrm{d}\beta}{\mathrm{d}s} = \frac{\mathrm{d}\gamma}{\mathrm{d}s} = 0,$$

whence it is clear that α, β, γ are constants. But we know from [I, 160] that the direction-cosines α, β, γ of the tangent are equal respectively to dx/ds, dy/ds, dz/ds, and since these are constant, the coordinates x, y, z must be first degree polynomials in s, i.e. we have in fact a straight line.

It may easily be shown in a similar way that *if the torsion is zero along a curve, the curve lies in a plane.*

126. The osculating plane. The plane defined by vectors **t** and **n** is called *the osculating plane to the curve*. The normal to this plane is given by **b**, the direction-cosines of which we proceed to find.

Since **b** is a unit vector, its direction-cosines are equal to its components b_x, b_y, b_z. It follows from (17) that:

$$\left.\begin{aligned} \alpha_2 &= b_x = t_y n_z - t_z n_y; \\ \beta_2 &= b_y = t_z n_x - t_x n_z; \\ \gamma_2 &= b_z = t_x n_y - t_y n_x, \end{aligned}\right\} \tag{29}$$

where $t_x, \ldots, n_x, \ldots$ are the components of **t** and **n**. As we saw above, t_x, t_y, t_z are proportional to dx, dy, dz; n_x, n_y, n_z are proportional to the components of **N**, which are equal to d^2x/ds^2, d^2y/ds^2, d^2z/ds^2, these latter being in turn, by (24), proportional to the differences

$$d^2x\,ds - d^2s\,dx, \quad d^2y\,ds - d^2s\,dy, \quad d^2z\,ds - d^2s\,dz. \tag{30}$$

We can thus replace t_x, t_y, t_z by dx, dy, dz in (29), n_x, n_y, n_z being replaced by differences (30). On cancelling, we find that the direction-cosines of the binormal are in fact proportional to the expressions:

$$\left.\begin{aligned} A = dy\,d^2z - dz\,d^2y; \quad B = dz\,d^2x - dx\,d^2z; \\ C = dx\,d^2y - dy\,d^2x, \end{aligned}\right\} \tag{31}$$

which we introduced above [124]. If we take (x, y, z) as the coordinates of a variable point M of the curve (L), we can write the equation of the osculating plane as

$$A(X - x) + B(Y - y) + C(Z - z) = 0.$$

At points where the length $|\mathbf{N}| = 0$, i.e. $\varrho = \infty$, all three of expressions (31) are zero, as follows from (27), and the osculating plane is not defined. The directions of the principal normal and binormal are also not defined.

127. The helix. Let (l) be the base in the XY plane of a cylinder with generators parallel to the z axis (Fig. 109). Let σ be the length of arc of (l) measured in a definite direction from its point of intersection A with OX, and let the equation of (l) be

$$x = \varphi(\sigma); \quad y = \psi(\sigma). \tag{32}$$

We mark off an arc AN on (l) and draw a straight line $NM = k\sigma$, parallel to the z axis, where k is a given numerical coefficient (the thread of the screw).

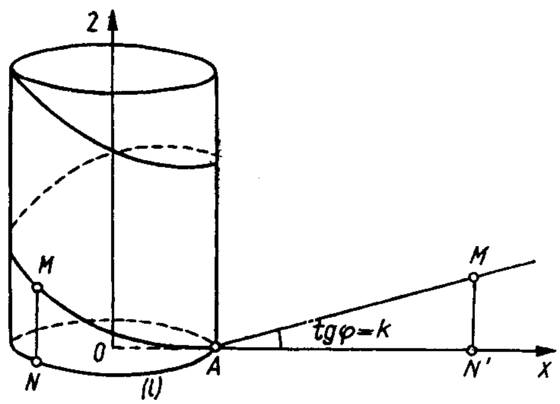

FIG. 109

The locus of the point M is a helix (L), inscribed on the cylinder. Its parametric equations are clearly:

$$x = \varphi(\sigma); \quad y = \varphi(\sigma); \quad z = k\sigma. \tag{33}$$

Let s be the length of arc of (L) measured from the point A. We have:

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = [\varphi'^2(\sigma) + \psi'^2(\sigma) + k^2]\,\mathrm{d}\sigma^2.$$

But $\varphi'(\sigma)$ and $\psi'(\sigma)$ are respectively equal to the cosine and sine of the angle formed by the tangent to (l) with the x axis [I, 70], whence $\varphi'^2(\sigma) + \psi'^2(\sigma) = 1$, and we can write the above expression as

$$\mathrm{d}s = \sqrt{1 + k^2}\,\mathrm{d}\sigma,$$

so that

$$s = \sqrt{1 + k^2}\,\sigma,$$

We now find the cosine of the angle formed by the tangent to (L) with the z axis:

$$\gamma = \frac{\mathrm{d}z}{\mathrm{d}s} = \frac{\mathrm{d}z}{\mathrm{d}\sigma} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}s} = \frac{k}{\sqrt{1 + k^2}};$$

this gives *the first property of a helix: the tangent to a helix forms a constant angle with a fixed direction.*

We recall the third of expressions (28), which gives in this case:

$$0 = \frac{\gamma_1}{\varrho} \quad \text{or} \quad \gamma_1 = 0,$$

so that the principal normal to a helix is perpendicular to the z axis, i.e. to generators of the cylinder. On the other hand, it is perpendicular to the tangent to the helix. A generator and a tangent are easily seen to define the tangent plane to the cylinder at the point of the helix concerned, and it follows from the above that the principal normal is perpendicular to this tangent plane. Hence we have *a second property of the helix: the principal normal at any point of a helix coincides with the normal to the cylinder on which it is traced.*

We now return to the cosines $\gamma, \gamma_1, \gamma_2$ of the angles formed by the z axis with the directions of a movable triad of the helix. On noting that $\gamma^2 + \gamma_1^2 + \gamma_2^2 = 1$ and that γ and γ_1 are constants, as we saw above, we can conclude that γ_2 is also constant. The third of expressions (28_2) gives in the present case $-(\gamma/\varrho) - (\gamma_2/\tau) = 0$, whence we see that the ratio ϱ/τ is constant; this gives us *the third property of a helix: the ratio of the radius of curvature to the radius of torsion is constant along a helix.* Let r denote the radius of curvature of the plane curve (l). In view of the square of the curvature being equal to the sum of the squares of the second derivatives of the coordinates with respect to the length of arc, we can write:

$$\frac{1}{r^2} = \varphi''^2(\sigma) + \psi''^2(\sigma)$$

and

$$\frac{1}{\varrho^2} = \left(\frac{d^2x}{ds^2}\right)^2 + \left(\frac{d^2y}{ds^2}\right)^2 + \left(\frac{d^2z}{ds^2}\right)^2 = \left[\left(\frac{d^2x}{d\sigma^2}\right)^2 + \left(\frac{d^2y}{d\sigma^2}\right)^2 + \left(\frac{d^2z}{d\sigma^2}\right)^2\right]\frac{1}{(1+k^2)^2},$$

whence

$$\frac{1}{\varrho^2} = \frac{\varphi''^2(\sigma)}{(1+k^2)^2} + \frac{\psi''^2(\sigma)}{(1+k^2)^2} = \frac{1}{(1+k^2)^2 r^2},$$

or $\varrho = (1 + k^2)\, r$, i.e. the radius of curvature of a helix differs only by a constant factor from the radius of curvature of the guide-curve at the corresponding point. If the cylinder is circular, i.e. the guide-curve (l) is a circle, r is constant, so that ϱ is also constant; further, by the third property above, τ is now likewise constant, i.e. *a helix on a circular cylinder has constant curvature and constant torsion.*

A further important property of a helix may be described in conclusion. If two points are taken on a cylinder, the shortest distance between them is given by the helix passing through the points. A helix on a cylinder is exactly analogous in this connection to a straight line on a plane. The present property is usually stated as: *the geodesics of a cylinder are helices. The geodesics of a surface are defined in general as the curves on the surface giving the shortest distance between two points.*

If we roll out the cylinder on the XZ plane by rotating it about a generator passing through the point A, the helix becomes a straight line on the plane since

the ratio of arc AN to the straight segment NM retains the constant value $1/k$. Lengths will be preserved in the unrolling, and the above property — that the helix gives the shortest distance between two points of the cylinder — becomes obvious. This property is in fact directly related to the second property of a helix, i.e. to the fact that the principal normals of a helix coincide with the normals to the cylinder. It may be shown generally in geometry that *the principal normals to the geodesics of any surface are identical with the normals to the surface.*

128. Field of unit vectors. Let $\mathbf{t}$ be a field of unit vectors, i. e. a unit vector $\mathbf{t}$ is given at every point of space. We deduce a simple and important expression for the curvature vector $\mathbf{N}$ of vector lines of the field. If coordinates (x, y, z) are taken, and s is the length of arc of a vector line, we can write:

$$\frac{\mathrm{d}x}{\mathrm{d}s} = t_x; \quad \frac{\mathrm{d}y}{\mathrm{d}s} = t_y; \quad \frac{\mathrm{d}z}{\mathrm{d}s} = t_z.$$

We find the component N_x of the curvature vector:

$$N_x = \frac{\mathrm{d}t_x}{\mathrm{d}s} = \frac{\partial t_x}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}s} + \frac{\partial t_x}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}s} + \frac{\partial t_x}{\partial z} \cdot \frac{\mathrm{d}z}{\mathrm{d}s}$$

or

$$N_x = \frac{\partial t_x}{\partial x} t_x + \frac{\partial t_x}{\partial y} t_y + \frac{\partial t_x}{\partial z} t_z.$$

On differentiating the equation

$$t_x^2 + t_y^2 + t_z^2 = 1$$

with respect to x, we get:

$$t_x \frac{\partial t_x}{\partial x} + t_y \frac{\partial t_y}{\partial x} + t_z \frac{\partial t_z}{\partial x} = 0.$$

By subtracting this sum from the above expression for N_x, we can write:

$$N_x = \left(\frac{\partial t_x}{\partial z} - \frac{\partial t_z}{\partial x}\right) t_z - \left(\frac{\partial t_y}{\partial x} - \frac{\partial t_x}{\partial y}\right) t_y,$$

i.e. $N_x = (\operatorname{curl} \mathbf{t} \times \mathbf{t})_x$, whilst similar expressions can clearly be written for the other two components. The required expression for the curvature vector of a vector line is thus:

$$\mathbf{N} = \operatorname{curl} \mathbf{t} \times \mathbf{t}. \tag{34}$$

The necessary and sufficient condition for these lines to be straight is that the length of $\mathbf{N}$, i.e. the curvature $1/\varrho$, is zero [125]. Hence *the necessary and sufficient condition for the vector lines of a unit field* $\mathbf{t}$ *to be straight is that:*

$$\operatorname{curl} \mathbf{t} \times \mathbf{t} = 0. \tag{35}$$

Furthermore, we saw in [110] that the necessary and sufficient condition for the existence of a family of surfaces orthogonal to the vector lines is:

$$\operatorname{curl} \mathbf{t} \cdot \mathbf{t} = 0. \tag{36}$$

Conditions (35) and (36) can only be satisfied simultaneously when curl $\mathbf{t} = 0$, since non-vanishing of this vector implies its being parallel to $\mathbf{t}$ by (35), whilst (36) means that the two vectors are perpendicular. Hence, *the vector lines of a unit field* $\mathbf{t}$ *are normal to a family of surfaces only when* curl $\mathbf{t} = 0$. This proposition plays an important part in explaining the principles of geometrical optics.

§ 13. Elementary theory of surfaces

129. The parametric equations of a surface. We have so far considered the equation of a surface in space with (x, y, z) axes in the explicit form $z = f(x, y)$ or implicitly as

$$F(x, y, z) = 0. \tag{37}$$

The equation of a surface may be written parametrically, the coordinates of its points being expressed as functions of two variable parameters u and v:

$$x = \varphi(u, v); \quad y = \psi(u, v); \quad z = \omega(u, v). \tag{38}$$

We shall assume that these functions are single-valued and continuous, and possess continuous derivatives up to the second order in a certain domain of variation of parameters (u, v).

If we substitute the expressions for the coordinates in terms of u and v in the left-hand side of (37), we must get an identity in u, v. We have on differentiating this identity with respect to the independent variables u, v:

$$\frac{\partial F}{\partial x} \cdot \frac{\partial \varphi}{\partial u} + \frac{\partial F}{\partial y} \cdot \frac{\partial \psi}{\partial u} + \frac{\partial F}{\partial z} \cdot \frac{\partial \omega}{\partial u} = 0,$$
$$\frac{\partial F}{\partial x} \cdot \frac{\partial \varphi}{\partial v} + \frac{\partial F}{\partial y} \cdot \frac{\partial \psi}{\partial v} + \frac{\partial F}{\partial z} \cdot \frac{\partial \omega}{\partial v} = 0.$$

If we consider these as two simultaneous equations in $\partial F/\partial x$, $\partial F/\partial y$, $\partial F/\partial z$ and use the algebraic lemma given in [104], we get:

$$\frac{\partial F}{\partial x} = k\left(\frac{\partial \psi}{\partial u} \cdot \frac{\partial \omega}{\partial v} - \frac{\partial \omega}{\partial u} \cdot \frac{\partial \psi}{\partial v}\right);$$
$$\frac{\partial F}{\partial y} = k\left(\frac{\partial \omega}{\partial u} \cdot \frac{\partial \varphi}{\partial v} - \frac{\partial \varphi}{\partial u} \cdot \frac{\partial \omega}{\partial v}\right);$$
$$\frac{\partial F}{\partial z} = k\left(\frac{\partial \varphi}{\partial u} \cdot \frac{\partial \psi}{\partial v} - \frac{\partial \psi}{\partial u} \cdot \frac{\partial \varphi}{\partial v}\right),$$

where k is a coefficient of proportionality.

We assume that k and at least one of the differences in brackets on the right-hand sides of the above expressions differ from zero.

We denote the three differences as follows, for brevity:

$$\frac{\partial\psi}{\partial u}\cdot\frac{\partial\omega}{\partial v}-\frac{\partial\omega}{\partial u}\cdot\frac{\partial\psi}{\partial v}=\frac{\mathrm{d}(y,z)}{\mathrm{d}(u,v)};$$

$$\frac{\partial\omega}{\partial u}\cdot\frac{\partial\varphi}{\partial v}-\frac{\partial\varphi}{\partial u}\cdot\frac{\partial\omega}{\partial v}=\frac{\mathrm{d}(z,x)}{\mathrm{d}(u,v)};$$

$$\frac{\partial\varphi}{\partial u}\cdot\frac{\partial\psi}{\partial v}-\frac{\partial\psi}{\partial u}\cdot\frac{\partial\varphi}{\partial v}=\frac{\mathrm{d}(x,y)}{\mathrm{d}(u,v)}.$$

We know from [I, 160] that the tangent plane at a point (x, y, z) of our surface can be written as:

$$\frac{\partial F}{\partial x}(X-x)+\frac{\partial F}{\partial y}(Y-y)+\frac{\partial F}{\partial z}(Z-z)=0,$$

or, if $\partial F/\partial x$, $\partial F/\partial y$, $\partial F/\partial z$ are replaced by their proportionate values, the tangent plane may be written:

$$\frac{\mathrm{d}(y,z)}{\mathrm{d}(u,v)}(X-x)+\frac{\mathrm{d}(z,x)}{\mathrm{d}(u,v)}(Y-y)+\frac{\mathrm{d}(x,y)}{\mathrm{d}(u,v)}(Z-z)=0. \quad (39)$$

The coefficients in this equation are known to be proportional to the direction-cosines of the normal to the surface.

The position of a variable point M of the surface is characterized by the values of the parameters u, v, and these are usually referred to as the coordinates of the point on the surface.

On assigning constant values to u and v, we obtain two families of lines on the surface: the so-called coordinate lines $u = C_1$, along which only v varies, and the coordinate lines $v = C_2$, along which only u varies. The two families give rise to a coordinate mesh on the surface.

Let us take the example of a sphere of radius R with centre at the origin. The parametric equations of the sphere may be written in the form:

$$x = R\sin u\cos v;\quad y = R\sin u\sin v;\quad z = R\cos u.$$

Coordinate lines $u = C_1$ and $v = C_2$ evidently consist here of parallels and meridians of the sphere.

We can characterize a surface without reference to coordinate axes by the radius vector $\mathbf{r}(u, v)$ drawn from a fixed point O to a variable

point M of the surface. The partial derivatives of this radius vector with respect to the parameters, $\mathbf{r}'_u$ and $\mathbf{r}'_v$, are evidently vectors directed along the tangents to the coordinate lines. The components of these, vectors on the axes OX, OY, OZ are, by (38), φ'_u, ψ'_u, ω'_u and φ'_v, ψ'_v, ω'_v and it is clear from this that the coefficients in the equation of the tangent plane (39) are the components of the vector product $\mathbf{r}'_u \times \mathbf{r}'_v$. This latter is perpendicular to the tangents $\mathbf{r}'_u$ and $\mathbf{r}'_v$ and is therefore directed along the normal to the surface; the square of its length is clearly given by the scalar product of $\mathbf{r}'_u \times \mathbf{r}'_v$ with itself, or to put the matter more simply, by the square of this vector product.† The unit vector along the normal to the surface plays an important part later, and we can evidently write this as:

$$\mathbf{m} = \frac{\mathbf{r}'_u \times \mathbf{r}'_v}{\sqrt{(\mathbf{r}'_u \times \mathbf{r}'_v)^2}}\,. \tag{40}$$

If we change the order of the factors in the vector product, we get the opposite direction for (40). We shall later fix a definite order for the factors, i.e. we shall define precisely the direction of the normal to the surface.

Let M be any point of the surface and let (L) be any curve lying on the surface and passing through the point M. The curve will not in general be a coordinate line, and both u and v will vary along it. The direction of the tangent to the curve will be given by the vector $\mathbf{r}'_u + \mathbf{r}'_v\, \mathrm{d}v/\mathrm{d}u$, if we assume that the parameter v is a function of u possessing a derivative in the neighbourhood of M along (L). It is clear from this that *the direction of the tangent at any point M of a curve lying on a surface is fully defined by the quantity* $\mathrm{d}v/\mathrm{d}u$ *at this point.* We assumed when defining the tangent plane and deducing its equation (39) that functions (38) have continuous partial derivatives at and in the neighbourhood of the point in question and that at least one of the coefficients of equation (39) differs from zero at the point.

If $\mathrm{d}(x, y)/\mathrm{d}(u, v) \neq 0$ for $u = u_0$, $v = v_0$, the same will be true in some neighbourhood of these values. This neighbourhood transforms, by the first two of expressions (38), into a neighbourhood of the values $x_0 = \varphi(u_0, v_0)$, $y_0 = \psi(u_0, v_0)$, and the first two of equations (38) can be solved with respect to u and v [I, 157] for (x, y) sufficiently close to (x_0, y_0), i.e. u and v can be expressed in terms of x and y.

† In general, we shall denote the square of the length of a vector $\mathbf{A}$ as $\mathbf{A}^2$, i.e. the scalar product $\mathbf{A} \cdot \mathbf{A}$.

Substitution of these expressions in the third of equations (38) gives the explicit equation of the surface $z = f(x, y)$ in the neighbourhood of the point concerned.

130. Gauss first differential form. We now consider the square of the differential of arc of any curve on a surface:

$$ds^2 = dx^2 + dy^2 + dz^2 =$$
$$= \left(\frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv\right)^2 + \left(\frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv\right)^2 + \left(\frac{\partial z}{\partial u} du + \frac{\partial z}{\partial v} dv\right)^2.$$

On removing the brackets, we get what is known as the *Gauss first differential form:*

$$ds^2 = E(u, v)\, du^2 + 2F(u, v)\, du\, dv + G(u, v)\, dv^2, \tag{41}$$

where

$$\left.\begin{aligned} E(u, v) &= \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2 \\ F(u, v) &= \frac{\partial x}{\partial u} \cdot \frac{\partial x}{\partial v} + \frac{\partial y}{\partial u} \cdot \frac{\partial y}{\partial v} + \frac{\partial z}{\partial u} \cdot \frac{\partial z}{\partial v} \\ G(u, v) &= \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2, \end{aligned}\right\} \tag{42}$$

or

$$E = \mathbf{r}_u'^2; \quad F = \mathbf{r}_u' \cdot \mathbf{r}_v'; \quad G = \mathbf{r}_v'^2. \tag{42$_1$}$$

We can show, exactly as in [119], that the vanishing of the coefficient F is the necessary and sufficient condition for the coordinate lines $u = C_1$ and $v = C_2$ to be mutually perpendicular. The curvilinear coordinates u, v on the surface are called orthogonal coordinates in this particular case.

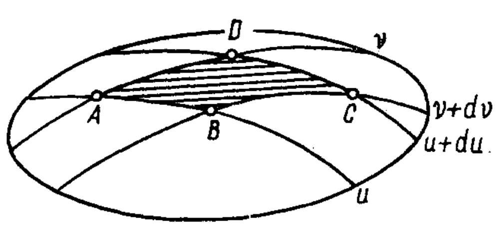

FIG. 110

We now find an expression for an elementary area of the surface in terms of the coefficients of (41). We take the small area bounded by two pairs of neighbouring coordinate lines (Fig. 110). Let (u, v) be the coordinates of vertex A. The sides AD and AB are respectively $\mathbf{r}_u'\, du$ and $\mathbf{r}_v'\, dv$. If we look on the small area as a parallelogram [cf. 57], its area can be written as the length of the vector product of these two vectors, i.e.

$$dS = |\mathbf{r}_u'\, du \times \mathbf{r}_v'\, dv| = |\mathbf{r}_u' \times \mathbf{r}_v'|\, du\, dv.$$

We have for the square of the length of the vector product:

$$(\mathbf{r}_u' \times \mathbf{r}_v')^2 = \left(\frac{\partial y}{\partial u}\cdot\frac{\partial z}{\partial v} - \frac{\partial z}{\partial u}\cdot\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial u}\cdot\frac{\partial x}{\partial v} - \frac{\partial x}{\partial u}\cdot\frac{\partial z}{\partial v}\right)^2 + \\ + \left(\frac{\partial x}{\partial u}\cdot\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\cdot\frac{\partial x}{\partial v}\right)^2,$$

whence, by identity (26) of [124]:

$$(\mathbf{r}_u' \times \mathbf{r}_v')^2 = EG - F^2, \tag{43}$$

so that we finally have for an elementary surface area:

$$\mathrm{d}S = \sqrt{EG - F^2}\,\mathrm{d}u\,\mathrm{d}v. \tag{44}$$

Furthermore, by substituting (43) in (40), we can write the expression for the unit normal to the surface as

$$\mathbf{m} = \frac{\mathbf{r}_u' \times \mathbf{r}_v'}{\sqrt{EG - F^2}}. \tag{45}$$

It may be pointed out that $EG - F^2$ is positive, by (43).

131. Gauss second differential form. We consider a curve (L) on a surface and let $\mathbf{t}$ be its unit tangential vector. This is obviously perpendicular to the unit normal vector to the surface, i.e. $\mathbf{t}\cdot\mathbf{m} = 0$. We have on differentiating this relationship with respect to the length of arc s of (L):

$$\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s}\cdot\mathbf{m} + \mathbf{t}\cdot\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}s} = 0 \quad \text{or} \quad \frac{1}{\varrho}(\mathbf{n}\cdot\mathbf{m}) + \mathbf{t}\cdot\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}s} = 0,$$

where ϱ is the radius of curvature and $\mathbf{n}$ the unit principal normal to L). The above equation can be rewritten as

$$\frac{\mathbf{n}\cdot\mathbf{m}}{\varrho} = -\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\cdot\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}s} \quad \text{or} \quad \frac{\cos\varphi}{\varrho} = -\frac{\mathrm{d}\mathbf{r}\cdot\mathrm{d}\mathbf{m}}{\mathrm{d}s^2},$$

where φ is the angle between the normal to the surface and the principal normal to (L). We can express the differentials $\mathrm{d}\mathbf{r}$ and $\mathrm{d}\mathbf{m}$ in terms of the coordinate parameters u and v and thus write:

$$\frac{\cos\varphi}{\varrho} = \frac{-(\mathbf{r}_u'\,\mathrm{d}u + \mathbf{r}_v'\,\mathrm{d}v)\cdot(\mathbf{m}_u'\,\mathrm{d}u + \mathbf{m}_v'\,\mathrm{d}v)}{\mathrm{d}s^2}. \tag{46}$$

On removing the brackets in the numerator, we get the *Gauss second differential form:*

$$-(\mathbf{r}_u'\,\mathrm{d}u + \mathbf{r}_v'\,\mathrm{d}v)\cdot(\mathbf{m}_u'\,\mathrm{d}u + \mathbf{m}_v'\,\mathrm{d}v) = \\ = L(u,v)\,\mathrm{d}u^2 + 2M(u,v)\,\mathrm{d}u\,\mathrm{d}v + N(u,v)\,\mathrm{d}v^2,$$

where

$$L = -\mathbf{r}'_u \cdot \mathbf{m}'_u; \quad M = -\frac{1}{2}(\mathbf{r}'_u \cdot \mathbf{m}'_v) - \frac{1}{2}(\mathbf{r}'_v \cdot \mathbf{m}'_u); \tag{47}$$
$$N = -\mathbf{r}'_v \cdot \mathbf{m}'_v,$$

so that (46) takes the final form:

$$\frac{\cos\varphi}{\varrho} = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2}\,. \tag{48}$$

We now notice alternative expressions for the coefficients L, M, and N. Differentiation of the obvious relationships

$$\mathbf{r}'_u \cdot \mathbf{m} = 0, \quad r'_v \cdot \mathbf{m} = 0$$

with respect to the independent variables u, v gives us the four equations:

$$\mathbf{r}''_{u^2} \cdot \mathbf{m} + \mathbf{r}'_u \cdot \mathbf{m}'_u = 0; \quad \mathbf{r}''_{uv} \cdot \mathbf{m} + \mathbf{r}'_u \cdot \mathbf{m}'_v = 0;$$
$$\mathbf{r}''_{vu} \cdot \mathbf{m} + \mathbf{r}'_v \cdot \mathbf{m}'_u = 0; \quad \mathbf{r}''_{v^2} \cdot \mathbf{m} + \mathbf{r}'_v \cdot \mathbf{m}'_v = 0,$$

and from these, together with (47), we can write the following expressions for the coefficients of the Gauss second differential form:

$$L = \mathbf{r}''_{u^2} \cdot \mathbf{m}; \quad N = \mathbf{r}''_{v^2} \cdot \mathbf{m}; \tag{49}$$
$$M = \mathbf{r}''_{uv} \cdot \mathbf{m} = -\mathbf{r}'_u \cdot \mathbf{m}'_v = -\mathbf{r}'_v \cdot \mathbf{m}'_u.$$

On recalling expression (45) for the vector $\mathbf{m}$, we can write (49) in the form:

$$L = \frac{\mathbf{r}''_{u^2} \cdot (\mathbf{r}'_u \times \mathbf{r}'_v)}{\sqrt{EG - F^2}}; \quad M = \frac{\mathbf{r}''_{uv} \cdot (\mathbf{r}'_u \times \mathbf{r}'_v)}{\sqrt{EG - F^2}};$$
$$N = \frac{\mathbf{r}''_{v^2} \cdot (\mathbf{r}'_u \times \mathbf{r}'_v)}{\sqrt{EG - F^2}}\,. \tag{50}$$

We now take the case when the equation of the surface is given explicitly:

$$z = f(x, y). \tag{51}$$

The role of parameters is now played by x and y, and we have the following expressions for the components of the radius vector and its derivatives with respect to the parameters:

$$\mathbf{r}(x, y, z); \quad \mathbf{r}'_x(1, 0, p); \quad \mathbf{r}'_y(0, 1, q)$$
$$\mathbf{r}''_{x^2}(0, 0, r); \quad \mathbf{r}''_{xy}(0, 0, s); \quad \mathbf{r}''_{y^2}(0, 0, t),$$

where

$$p = \frac{\partial f}{\partial x}; \quad q = \frac{\partial f}{\partial y}; \quad r = \frac{\partial^2 f}{\partial x^2}; \quad s = \frac{\partial^2 f}{\partial x \, \partial y}; \quad t = \frac{\partial^2 f}{\partial y^2}. \tag{52}$$

Use of (42_1) and (50) gives us the coefficients of both Gauss forms as:

$$F = 1 + p^2; \quad F = pq; \quad G = 1 + q^2;$$

$$L = \frac{r}{\sqrt{1 + p^2 + q^2}}; \quad M = \frac{s}{\sqrt{1 + p^2 + q^2}}; \quad N = \frac{t}{\sqrt{1 + p^2 + q^2}}. \tag{53}$$

We now make a definite choice of axes by taking the origin at a point M_0 of the surface, OX and OY in the tangent plane, and OZ along the normal to the surface at M_0. We use the zero subscript to indicate that a magnitude is being taken at M_0. With the present choice of axes, the cosines of the angles formed by the normal to the surface with OX and OY at M_0 will be zero, so that we have [62] $p_0 = q_0 = 0$, whilst (53) gives at M_0:

$$L_0 = r_0; \quad M_0 = s_0; \quad N_0 = t_0. \tag{54}$$

132. The curvature of lines ruled on surfaces. We return to (48). Its right-hand side depends on the values of the coefficients of the two Gauss forms and on the ratio dv/du. The last statement is immediately clear on dividing numerator and denominator by du^2. The coefficients are functions of the parameters u, v and have a definite numerical value at a given point of the surface. As we saw in [129], the ratio dv/du characterizes the direction of the tangent to the curve concerned. We can therefore say that both sides of (48) must have a definite value if we fix the point on the surface and the direction of the tangent to the curve on the surface. If we take two curves through a fixed point on the surface with the same tangential direction and the same principal normal, the angle φ will be the same for both curves, and therefore, by (48), ϱ will also be the same. We thus have the following theorem:

First Theorem. *Two curves on a surface with the same tangent and principal normal at a given point have the same radius of curvature at this point.*

If an arbitrary line (L) is drawn on a surface and passes through a point M, the plane containing the tangent and principal normal to (L) at M will cut the surface in a plane curve (L_0) having the same tangent and principal normal as (L) and hence having the

same radius of curvature. The theorem thus proved enables the investigation of the curvature of any curve on a surface to be reduced to the study of the curvature of a plane section through the surface.

A normal section of a surface at a given point M is defined as the section by any plane passing through the normal at M. There is obviously an infinite set of normal sections, one particular section being specified by assigning a definite tangential direction in the tangent plane to the surface, i.e. we fix the value of dv/du. The principal normal to a normal section must be equal or opposite to the vector $\mathbf{m}$, so that the angle φ equals 0 or π, whence $\cos\varphi = \pm 1$.

Let (L) be any curve on a surface through a point M. The normal section corresponding to (L) at M is defined as the normal section having a tangent in common with (L) at M. Let ϱ be the radius of curvature of (L) and R the radius of curvature of the corresponding normal section. We confine our attention here to the point M. Since both curves have the same tangent, the right-hand side of (48) is the same in both cases, and we can write

$$\frac{\cos\varphi}{\varrho} = \frac{\pm 1}{R}, \quad \text{i.e.} \quad \varrho = \pm R \cdot \cos\varphi, \tag{55}$$

where φ is the angle between the principal normal to the curve and the normal to the surface. The last formula expresses the following theorem:

SECOND THEOREM (Meusnier's theorem). *The radius of curvature at any given point of a curve on a surface is equal to the product of the radius of curvature of the corresponding normal section at the point with the cosine of the angle between the normal to the surface and the principal normal to the curve.* An alternative statement of the theorem is: the radius of curvature of any curve on a surface is equal to the projection of the radius of curvature of the corresponding normal section (marked off along the normal to the surface) on the principal normal to the curve.

In the case of a sphere, a normal section is a great circle, and if we take (L) as any circle traced on the sphere, (55) reduces to the obvious relationship between the radii of the two circles (Fig. 111).

By the second theorem, investigation of the curvature of a curve on a surface reduces to investigation of the curvature of the normal section at a given point of the surface. We have seen that, for a normal section, we must take $\cos\varphi = \pm 1$ in (48). We shall agree to refer the $(-)$ sign to ϱ when it occurs, i.e. we shall take the radius

of curvature of a normal section as negative if the principal normal to the section is in the opposite direction to **m**, i.e. opposite to the chosen normal direction to the surface. With this agreement, the formula is valid for normal sections:

$$\frac{1}{R} = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2}\,. \tag{56}$$

It may again be recalled that the coefficients of the differential forms on the right of this expression have definite values, since we have fixed the point on the surface and the value of $1/R$ depends only on the ratio dv/du, i.e. on the choice of tangential direction. The denominator on the right of (56) always has a positive value, since it is the expression for ds^2, and the sign of the curvature $1/R$ of a normal section is therefore defined by the sign of the numerator. We may distinguish the following three cases:

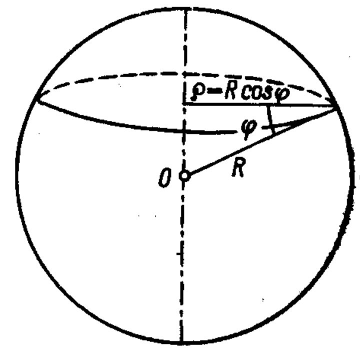

FIG. 111

1. If $M^2 - LN < 0$ at the point taken, $1/R$ has the same sign for all normal sections, i.e. the principal normals to all normal sections are directed towards the same side. Such a point on a surface is described as *elliptic*.

2. If $M^2 - LN > 0$, $1/R$ has different signs, i.e. there are normal sections at the point taken with opposite directions for the principal normal. A point of this kind is called *hyperbolic*.

3. If $M^2 - LN = 0$, the numerator of the right-hand side of (56) is a perfect square and $1/R$ does not change sign here, although it vanishes for one particular normal section. This kind of point is called *parabolic*.

It may be noted that the numerator of the right-hand side of (56) vanishes whilst changing sign in the hyperbolic case, and there are two normal sections with zero curvature. There are no such sections in the elliptic case.

We take axes as in [131], with the origin at the point of the surface taken and OX, OY situated in the tangent plane.

By (54), equation (56) takes the form:

$$\frac{1}{R} = \frac{r_0\,dx^2 + 2s_0\,dx\,dy + t_0\,dy^2}{ds^2}\,.$$

The tangent to the normal section lies in the XY plane, and the ratios dx/ds, dy/ds are equal respectively to $\cos\theta$ and $\sin\theta$, where θ is the angle formed by the tangent with the x axis. The above expression thus becomes:

$$\frac{1}{R} = r_0 \cos^2\theta + 2s_0 \cos\theta \sin\theta + t_0 \sin^2\theta. \tag{57}$$

This is an explicit expression of the dependence of the curvature $1/R$ on the direction of the tangent, characterised by the angle θ. Now, the point will be elliptic of $s_0^2 - r_0 t_0 < 0$, hyperbolic if $s_0^2 - - r_0 t_0 > 0$, and parabolic if $s_0^2 - r_0 t_0 = 0$.

In the case $s_0^2 - r_0 t_0 < 0$, the function $z = f(x, y)$ will have a zero maximum or minimum at the point concerned [I, 163], i.e. the surface near the point is situated on one side of the tangent plane. With $s_0^2 - r_0 t_0 \ 0, >$ there is neither a maximum nor minimum, i.e. the surface is situated on both sides of the tangent plane in any neighbourhood of the point. Finally, at a parabolic point where $s_0^2 - r_0 t_0 = 0$, nothing definite can be said of the disposition of the surface relative to the tangent plane.

It follows directly from (53) that the sign of $(M^2 - LN)$ is the same as that of $(s^2 - rt)$ for any choice of XYZ axes, so that the point is elliptic for $s^2 - rt < 0$, hyperbolic for $s^2 - rt > 0$, and parabolic for $s^2 - rt = 0$.

The same surface may have different kinds of point. For instance, in the case of a torus, obtained by rotation of a circle about an axis lying outside it but in the same plane [I, 107], points lying on the outward side are elliptic whilst points on the inward side are hyperbolic. These domains are separated from each other by the extreme parallels of the torus, all the points of which are parabolic.

133. Dupin's indicatrix and Euler's formula. Having fixed the coordinate axes as in the previous article, we draw an auxiliary curve in the tangent plane, i.e. the XY plane, as follows: we mark on each radius vector from the origin O a length $ON = \sqrt{\pm R}$, where R is the radius of curvature of the normal section for which the radius vector taken is a tangent. The $(\pm)$ sign is taken so that the quantity under the radical is positive. The locus of the point N is a curve called the *Dupin indicatrix.* The curve has the following property by construction: the square of the radius vector to any

point of it gives the absolute value of the radius of curvature of the normal section of which the radius vector is a tangent (Fig. 112).

The equation of the indicatrix is obtained as follows: let (ξ, η) be the coordinates of any point N of it. We have by construction:

$$\xi = \sqrt{\pm R}\cos\theta;$$

$$\mu = \sqrt{\pm R}\sin\theta,$$

i.e.

$$\xi^2 = \pm R\cos^2\theta;$$

$$\mu^2 = \pm R\sin^2\theta,$$

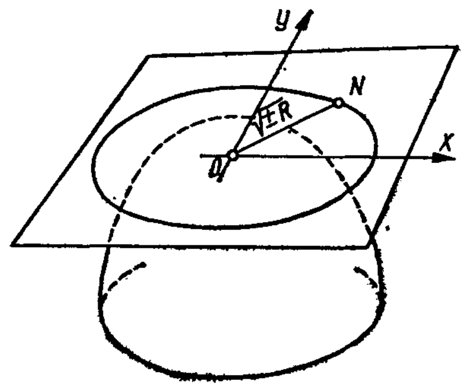

FIG. 112

where the upper sign must be taken for positive R, and the lower sign for negative R. Multiplication of both sides of (57) by $\pm R$ obviously gives us:

$$r_0\,\xi^2 + 2s_0\,\xi\eta + t_0\,\eta^2 = \pm 1. \tag{58}$$

This is the equation of the indicatrix. The curve gives a geometrical picture of the change in radius of curvature as a normal section rotates about a normal to the surface. In the case of an elliptic point, (58) gives an ellipse and a definite sign has to be taken on the right. We get two conjugate hyperbolas from (58) in the case of a hyperbolic point; with a parabolic point, the left-hand side is a perfect square and (58) may be re-written as:

$$k(a\xi + b\eta)^2 = \pm 1, \text{ i.e. } (a\xi + b\eta)^2 = \pm\frac{1}{k} = l^2$$

or

$$a\xi + b\eta = \pm l,$$

giving a set of two parallel straight lines. The curve has its centre at O in all three cases and has two axes of symmetry. We can take these as the x and y axes, in which case we know that the term in $\xi\eta$ falls out on the left-hand side of (58), i.e. we must have $s_0 = 0$, so that (57) now becomes:

$$\frac{1}{R} = r_0\cos^2\theta + t_0\sin^2\theta. \tag{59}$$

The geometrical significance of coefficients r_0 and t_0 may be seen as follows. On setting $\theta = 0$ in (59), we get the curvature $1/R_1$ of the normal section to which the x axis is the tangent, and we have $r_0 = 1/R_1$. Similarly, on setting $\theta = \pi/2$, we get $t_0 = 1/R_2$, where $1/R_2$ is the curvature of the normal section to which OY is tangential. We get Euler's formula on substituting the values found for r_0 and t_0 in (59):

$$\frac{1}{R} = \frac{\cos^2\theta}{R_1} + \frac{\sin^2\theta}{R_2}. \tag{60}$$

It will be recalled that our x and y axes are the axes of symmetry of curve (58). We assume that $1/R_1 \neq 1/R_2$ and that, for instance, $1/R_1 > 1/R_2$. It follows at once from (60) that $1/R$ attains its greatest value at $\theta = 0$ and $\theta = \pi$, and its least value at $\theta = \pi/2$ and $\theta = 3\pi/2$.

The result obtained may be stated as the following theorem:

THIRD THEOREM. *There exist at any point of a surface two mutually perpendicular directions lying in the tangent plane for which the curvature* $1/R$ *has a maximum and minimum; if the curvatures corresponding to these directions are* $1/R_1$ *and* $1/R_2$*, the curvature of any normal section is given by* (60), *where* θ *is the angle that the tangent to the normal section forms with the direction that gives* $1/R_1$.

We refer to R_1 and R_2 as the *principal radii of curvature of normal sections* at the point concerned. The two directions in the tangent plane that give rise to these are called the *principal directions.* Moreover, in the hyperbolic case it is useful to distinguish two further directions in the tangent plane, those of the asymptotes to the indicatrix. The radius vectors to the indicatrix in these *asymptotic directions* are infinite, and the curvatures of the corresponding normal sections are zero at the point taken.

In the elliptic case, R_1 and R_2 have the same sign, whilst their signs are opposite in the hyperbolic case. The curvature of one of the principal normal sections becomes zero in the parabolic case; if we take say $1/R_2 = 0$, we get the formula for the parabolic case:

$$\frac{1}{R} = \frac{\cos^2\theta}{R_1}.$$

A particular case of elliptic points on a surface may be noticed, when R_1 and R_2 are equal. We have in this case from (60): $1/R = 1/R_1$, i.e. all normal sections have the same curvature at the point taken. A point of this sort on a surface is called an *umbilic*. A surface becomes very like a sphere in the neighbourhood of an umbilic. It can be shown that a sphere is the only surface, all the points of which are umbilics.

134. Finding the principal radii of curvature and principal directions. The basic expression (56) for the curvature of a normal section may be written in the form

$$(L - ER^{-1})\,\mathrm{d}u^2 + 2(M - FR^{-1})\,\mathrm{d}u\,\mathrm{d}v + (N - GR^{-1})\,\mathrm{d}v^2 = 0. \quad (61)$$

On dividing by $\mathrm{d}v^2$ and bringing in the auxiliary $t = \mathrm{d}u/\mathrm{d}v$, characterizing the tangential direction to the normal section, we get the equation:

$$\varphi(R^{-1}, t) = (L - ER^{-1})\,t^2 + 2(M - FR^{-1})\,t + (N - GR^{-1}) = 0,$$

which gives the curvature R^{-1} of the normal section as a function of t. The value of R^{-1} must be a maximum or minimum for the principal directions, so that the derivative of R^{-1} with respect to t must vanish. But this derivative is obviously given by [I, 69]:

$$\frac{\mathrm{d}R^{-1}}{\mathrm{d}t} = -\frac{\dfrac{\mathrm{d}\varphi}{\mathrm{d}t}}{\dfrac{\mathrm{d}\varphi}{\mathrm{d}R^{-1}}},$$

so that the derivative $\mathrm{d}\varphi/\mathrm{d}t$ must vanish for the principal directions, i.e.

$$\frac{1}{2}\,\frac{\mathrm{d}\varphi}{\mathrm{d}t} = (L - ER^{-1})\,t + (M - FR^{-1}) = 0.$$

On replacing $t = \mathrm{d}u/\mathrm{d}v$ and multiplying by $\mathrm{d}v$, we get:

$$(L - ER^{-1})\,\mathrm{d}u + (M - FR^{-1})\,\mathrm{d}v = 0. \quad (62)$$

If we were to divide (61) by $\mathrm{d}u^2$ and take $t_1 = \mathrm{d}v/\mathrm{d}u$ as the variable characterizing the direction of the tangent, we should arrive in the same way at the equation for the principal directions:

$$(M - FR^{-1})\,\mathrm{d}u + (N - GR^{-1})\,\mathrm{d}v = 0. \quad (63)$$

On taking the term in $\mathrm{d}v$ to the right-hand side in (62) and (63) and dividing the respective sides of the equations into each other, we obtain a quadratic equation for the curvatures $1/R_1$ and $1/R_2$ of the principal normal sections:

$$(EG - F^2)\,\frac{1}{R^2} + (2FM - EN - GL)\,\frac{1}{R} + (LN - M^2) = 0. \quad (64)$$

The quantity:

$$K = \frac{1}{R_1 R_2} \quad (65)$$

is called the *Gaussian curvature* of the surface at the given point, whilst the quantity

$$H = \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_1}\right) \tag{66}$$

is known as the mean curvature. We get directly from the quadratic equation (64) expressions for the Gaussian and mean curvatures in terms of the coefficients of the first and second Gauss forms:

$$K = \frac{LN - M^2}{EG - F^2}; \quad H = \frac{EN - 2FM + GL}{2(EG - F^2)}. \tag{67}$$

We now write equations (62) and (63) as

$$(L\,\mathrm{d}u + M\,\mathrm{d}v)R = E\,\mathrm{d}u + F\,\mathrm{d}v;$$
$$(M\,\mathrm{d}u + N\,\mathrm{d}v)\,R = F\,\mathrm{d}u + G\,\mathrm{d}v.$$

We eliminate R by dividing these equations into each other, and obtain after simple rearrangement:

$$(EM - FL)\,\mathrm{d}u^2 + (EN - GL)\,\mathrm{d}u\,\mathrm{d}v + (FN - GM)\,\mathrm{d}v^2 = 0. \tag{68}$$

Division by $\mathrm{d}u^2$ gives us a quadratic equation in $\mathrm{d}v/\mathrm{d}u$, the roots of which characterize the principal directions at a point of the surface:

$$\frac{\mathrm{d}v}{\mathrm{d}u} = \varphi_1(u, v); \quad \frac{\mathrm{d}v}{\mathrm{d}u} = \varphi_2(u, v). \tag{69}$$

135. Line of curvature. *A line of curvature is defined as a line on a surface such that the tangent at every point is along a principal direction.* Since there are two principal directions at every point of a surface, there will be two families of lines of curvature on the surface and the families will be mutually orthogonal. The aggregate of all lines of curvature thus gives rise to an orthogonal mesh on the surface. Equation (68) or the equivalent equations (69) represent differential equations for the lines of curvature; integration gives us v in terms of u, and substitution of the expression obtained in the equation of the surface leads to the actual equations of the lines of curvature.

Let us consider the conditions under which a given coordinate mesh on a surface represents a mesh of lines of curvature. First of all, the mesh must be orthogonal if it consists of lines of curvature, i.e. we must have $F = 0$. Furthermore, if the coordinate lines $u = C_1$ and $v = C_2$ are to be lines of curvature, equation (68) must be

satisfied on substituting constants for u and v. If we take into account the result already obtained, that $F = 0$, we have $GM = 0$ and $EM = 0$. But we have seen that $EG - F^2$ is positive, so that E and G cannot be zero; it follows from the above that we must have $M = 0$. Hence a necessary condition for the coordinate mesh to be a mesh of lines of curvature is that $F = M = 0$. Conversely, if this condition is satisfied, the differential equation (68) of the lines of curvature has the solution $u = C_1$ and $v = C_2$, i.e. the coordinate lines are lines of curvature; hence we have the following theorem: *a necessary and sufficient condition for a coordinate mesh to be a mesh of lines of curvature consists in the vanishing of the middle term in both the Gauss differential forms everywhere on the surface,* i.e. $F = M = 0$.

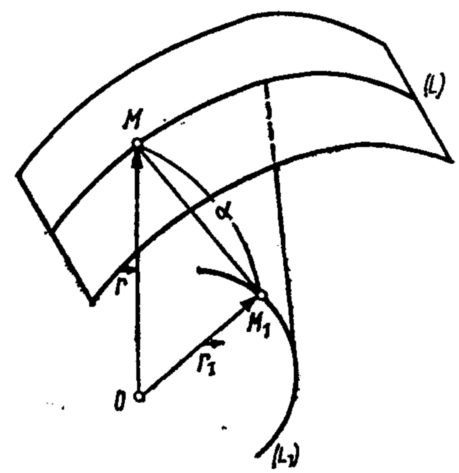

FIG. 113

It is possible to give a different definition of line of curvature to that at the beginning of the present article. Let (L) be a curve on a surface. The normals to the surface along (L) form a family of straight lines with a single parameter defining the position of the point on (L), and the family will not in general have an envelope. An envelope will exist, however, if the curve (L) is chosen in a particular way.† The conditions for a suitable choice will now be explained.

Let the curve (L) be chosen on the surface so that there exists an envelope (L_1) of normals to the surface along (L) (Fig. 113). Let $\mathbf{r}$ denote the radius vector to a point of (L), $\mathbf{r}_1$ the corresponding radius vector to a point of (L_1), and a the algebraic length of the normal measured from (L) to (L_1); then we can clearly write:

$$\mathbf{r}_1 = \mathbf{r} + a\mathbf{m}, \tag{70}$$

where $\mathbf{m}$ is as usual the unit normal to the surface. Since (L_1) is the envelope of the normals, the vector $d\mathbf{r}_1$, directed along its tangent,

† A family of straight lines in space, containing a single parameter, in general has no envelope, i.e. the lines are not tangents to any one curve. There is an envelope only in particular cases.

must be parallel to $\mathbf{m}$, and we can write $d\mathbf{r}_1 = b\mathbf{m}$, where b is a scalar. We get on differentiating (70):

$$b\mathbf{m} = d\mathbf{r} + a\, d\mathbf{m} + da\, \mathbf{m}, \quad \text{i.e.} \quad d\mathbf{r} + a\, d\mathbf{m} = c\mathbf{m}, \tag{71}$$

where c is a scalar. We show that $c = 0$. We form the scalar product of both sides of (71) with $\mathbf{m}$:

$$d\mathbf{r} \cdot \mathbf{m} + a\, d\mathbf{m} \cdot \mathbf{m} = c.$$

Vector $d\mathbf{r}$ is directed along the tangent to (L), i.e. is perpendicular to $\mathbf{m}$, so that $d\mathbf{r} \cdot \mathbf{m} = 0$. Moreover, it follows as usual from the equation $\mathbf{m} \cdot \mathbf{m} = 1$ that $d\mathbf{m} \cdot \mathbf{m} = 0$, and the above equation therefore in fact gives $c = 0$, whilst (71) becomes:

$$d\mathbf{r} + a\, d\mathbf{m} = 0. \tag{72}$$

This expression is generally known as Rodrigues' formula. It has been obtained from the assumption that normals to the surface along (L) have an envelope. We now assume the converse, that expression (72) is valid along a curve (L) on the surface. Formula (70) now defines a curve (L_1); on differentiating the formula and taking into account (72), we get $d\mathbf{r}_1 = da\mathbf{m}$, i.e. the direction of $\mathbf{m}$ and the tangent to (L_1) are parallel. In other words, normals to the surface along (L) are tangents to (L_1). Hence (72) gives the necessary and sufficient condition for the existence of an envelope of normals to the surface along (L). It must be noted that the envelope can degenerate to a point; the normals in this case form a cylindrical or conical surface, where it may be shown that condition (72) must still be fulfilled.

We write (72) in the expanded form:

$$\mathbf{r}_u' du + \mathbf{r}_v' dv + a(\mathbf{m}_u' du + \mathbf{m}_v' dv) = 0$$

and form the scalar product with $\mathbf{r}_u'$.

We get by (42_1), (47) and (49):

$$E\, du + F\, dv + a(-L du - M\, dv) = 0,$$

which is the same as equation (62) with $a = R$. Similarly, on forming the scalar product with $\mathbf{r}_v'$, we obtain equation (63). It is easy to show the converse, that equation (72) is obtained with $a = R$ from (62) and (63), which define the principal radii of curvature and principal directions. We shall not dwell on the proof. Condition (72) for the existence of a normal envelope is thus equivalent to (62) and (63),

a being one of the principal radii of curvature. These remarks lead us to the following result: *lines of curvature on a surface are characterized by the property that the normals to the surface along them have envelopes (or form cones or cylinders), the length of the normal between the surface and envelope being equal to one of the principal radii of curvature.*

If a plane curve rotates about an axis in its plane, the lines of curvature of the resulting surface of revolution are its meridians and parallels. The normals to the surface in fact form a plane along a meridian, and a cone along a parallel.

136. Dupin's theorem. Let three families of mutually orthogonal surfaces in space be:

$$\varphi(x, y, z) = q_1; \quad \psi(x, y, z) = q_2; \quad \omega(x, y, z) = q_3.$$

They form a mesh of orthogonal curvilinear coordinates in space [119]. The radius vector $\mathbf{r}$ from the origin to a variable point M in space is characterized by the curvilinear coordinates q_1, q_2, q_3 of the point. The partial derivatives $\mathbf{r}'_{q_1}, \mathbf{r}'_{q_2}, \mathbf{r}'_{q_3}$ give vectors directed along the tangents to coordinate lines, and since the coordinates are orthogonal we can write the vector equations:

$$\mathbf{r}'_{q_2} \cdot \mathbf{r}'_{q_3} = 0; \quad \mathbf{r}'_{q_3} \cdot \mathbf{r}'_{q_1} = 0; \quad \mathbf{r}'_{q_1} \cdot \mathbf{r}'_{q_2} = 0. \tag{73}$$

We differentiate the first of these equations with respect to q_1, the second with respect to q_2, and the third with respect to q_3:

$$\begin{gathered} \mathbf{r}''_{q_1q_2} \cdot \mathbf{r}'_{q_3} + \mathbf{r}'_{q_2} \cdot \mathbf{r}''_{q_1q_3} = 0 \\ \mathbf{r}''_{q_2q_3} \cdot \mathbf{r}'_{q_1} + \mathbf{r}'_{q_3} \cdot \mathbf{r}''_{q_1q_2} = 0 \\ \mathbf{r}''_{q_1q_3} \cdot \mathbf{r}'_{q_2} + \mathbf{r}'_{q_1} \cdot \mathbf{r}''_{q_2q_3} = 0. \end{gathered}$$

From these we obtain at once:

$$\mathbf{r}''_{q_1q_2} \cdot \mathbf{r}'_{q_3} = \mathbf{r}''_{q_2q_3} \cdot \mathbf{r}'_{q_1} = \mathbf{r}''_{q_3q_1} \cdot \mathbf{r}'_{q_2} = 0.$$

We take together the three equations:

$$\mathbf{r}'_{q_1} \cdot \mathbf{r}'_{q_3} = \mathbf{r}'_{q_2} \cdot \mathbf{r}'_{q_3} = \mathbf{r}''_{q_1q_2} \cdot \mathbf{r}'_{q_3} = 0.$$

It follows from these that the vectors $\mathbf{r}'_{q_1}, \mathbf{r}'_{q_2}$, and $\mathbf{r}''_{q_1q_2}$ are all perpendicular to the same vector $\mathbf{r}'_{q_3}$ and are therefore coplanar, whence it follows that [105]

$$\mathbf{r}''_{q_1q_2} \cdot (\mathbf{r}'_{q_1} \times \mathbf{r}'_{q_2}) = 0. \tag{74}$$

We now take the coordinate surface $q_3 = C$. Parameters q_1 and q_2 are coordinate parameters on it, and the coordinate lines $q_1 = C$ and $q_2 = C$ are the lines of its intersection with two other coordinate surfaces of our orthogonal coordinate set in space. We had the following expressions in [130, 131]:

$$F = \mathbf{r}'_{q_1} \cdot \mathbf{r}'_{q_2}; \quad M = \frac{\mathbf{r}''_{q_1 q_2} \cdot (\mathbf{r}'_{q_1} \times \mathbf{r}'_{q_2})}{\sqrt{EG - F^2}},$$

and equations (73) and (74) show that here $F = M = 0$, i.e. the q_1 and q_2 coordinate lines are lines of curvature on the surface $q_3 =$ $=$ const. This leads us to Dupin's theorem: *given three families of mutually orthogonal surfaces in space, any two surfaces of different families intersect in a line which is a line of curvature for both the surfaces.*

137. Examples. 1. The equation of the oblate ellipsoid of revolution:

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{c^2} = 1 \qquad (a^2 > c^2)$$

can be written in the parametric form:

$$x = a \cos u \sin v; \quad y = a \sin u \sin v; \quad z = c \cos v.$$

The coordinate lines $u = c_1$ are clearly the lines of intersection of the ellipsoid with the planes $y = x \tan c_1$, passing through the axis of rotation, i.e. they are meridians, whilst the coordinate lines $v = c_2$ are parallels, obtained by the intersection of the ellipsoid with the planes $z = c \cos c_2$, perpendicular to the axis of rotation. On applying formulae (42) and (50) of [130, 131] and taking into account the fact that x, y, z are the components of $\mathbf{r}$, we get:

$$E = a^2 \sin^2 v; \quad F = 0; \quad G = a^2 \cos^2 v + c^2 \sin v;$$

$$L = \frac{ac \sin^2 v}{\sqrt{a^2 \cos^2 v + c^2 \sin^2 v}}; \qquad M = 0; \qquad N = \frac{ac}{\sqrt{a^2 \cos^2 v + c^2 \sin^2 v}}.$$

The equation $F = M = 0$ may be foreseen from the fact that the meridians and parallels are lines of curvature of the ellipsoid of revolution. The remaining coefficients depend only on the parameter v, characterizing the position of a point on a meridian. The principal directions are clearly given by the tangents to meridians and parallels. The expression $(LN - M^2)$ is here positive over the entire surface, i.e. every point of the surface is elliptic. We note the Gaussian curvature, without working out the principal radii of curvature separately:

$$K = \frac{1}{R_1 R_2} = \frac{LN - M^2}{EG - F^2} = \frac{c^2}{(a^2 \cos v + c^2 \sin^2 v)^2}.$$

2. The equation of a cone of the second order:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$$

may be written explicitly as

$$z=c\sqrt{\frac{x^2}{a^2}+\frac{y^2}{b^2}}\,.$$

We easily find by direct differentiation that:

$$p=\frac{c^2\,x}{a^2\,z}\,;\qquad q=\frac{c^2\,y}{b^2\,z}\,;\qquad r=\frac{c^4\,y^2}{a^2\,b^2\,z^3}\,;$$

$$s=-\frac{c^4\,xy}{a^2\,b^2\,z^3}\,;\quad t=\frac{c^4\,x^2}{a^2\,b^2\,z^3}\,.$$

On using expressions (53), we can find all the coefficients of the Gauss forms. We remark simply that here $rt-s^2=0$, i.e. every point of the surface is parabolic, and one of the principal radii of curvature is infinite. The corresponding principal direction is evidently along the straight generator of the cone.

3. We consider the hyperbolic paraboloid

$$z=\frac{x^2}{2a^2}-\frac{y^2}{2b^2}\,.$$

Here, $r=1/a^2$, $s=0$ and $t=-1/b^2$, so that $rt-s^2<0$, and every point of the surface is therefore hyperbolic. The two straight generators of the surface in this case give the asymptotes of the indicatrix, which consists of two conjugate hyperbolas. The situation is similar in the case of a hyperboloid of one sheet.

4. Ordinary Cartesian, along with spherical and cylindrical, coordinates provide the simplest examples of orthogonal coordinates in space. A further example of such coordinates may be indicated. We take an equation of a second order surface containing a parameter ϱ:

$$\frac{x^2}{a^2+\varrho}+\frac{y^2}{b^2+\varrho}+\frac{z^2}{c^2+\varrho}-1=0, \tag{75}$$

where $a^2>b^2>c^2$. On fixing a point $M(x, y, z)$ and getting rid of the denominators, we arrive at a third degree equation in ϱ. It can be shown that this equation has three real roots u, v, w, contained respectively between the limits

$$+\infty>u>-c^2;\quad -c^2>v>-b^2;\quad -b^2>w>-a^2. \tag{76}$$

In fact, the left-hand side of equation (75) approaches (-1) for large positive values of ϱ, and has the $(-)$ sign, whilst for ϱ somewhat greater than $(-c^2)$, the term $z^2/(c^2+\varrho)$ has a large positive value and the left-hand side of (75) has the $(+)$ sign. There must therefore be a value of ϱ in the interval $(-c^2, \infty)$ for which the left-hand side of (75) vanishes. Similar reasoning shows the existence of roots in the intervals $(-b^2, -c^2)$ and $(-a^2, -b^2)$. The three numbers u, v, w are called the *elliptic coordinates* of the given point $M(x, y, z)$. Our discussion has assumed the non-vanishing of all three coordinates of the point (x, y, z). If this is not the case, an equation of lower degree than the third is obtained

for ϱ. If, say, $z = 0$, whilst x and y differ from zero, equation (75) will give u and v, whilst w has to be taken equal to $(-c^2)$.

We now investigate the coordinate surfaces in the elliptic system. On substituting $\varrho = u$ in equation (75), where u belongs to the interval $(-c^2, \infty)$, we get the surface:

$$\frac{x^2}{a^2+u} + \frac{y^2}{b^2+u} + \frac{z^2}{c^2+u} = 1, \tag{77}$$

which is evidently an *ellipsoid*, since all three denominators in (77) are positive by the first of inequalities (76). On substituting $\varrho = v$, where v belongs to $(-b^2, -c^2)$, we get the hyperboloid of one sheet:

$$\frac{x^2}{a^2+v} + \frac{y^2}{b^2+v} + \frac{z^2}{c^2+v} = 1, \tag{78}$$

since here $a^2 + v > b^2 + v > 0$ and $c^2 + v < 0$. Finally, substitution of $\varrho = = w$, where w belongs to $(-a^2, -b^2)$, gives the *hyperboloid of two sheets:*

$$\frac{x^2}{a^2+w} + \frac{y^2}{b^2+w} + \frac{z^2}{c^2+w} = 1. \tag{79}$$

We prove that the three coordinate surfaces are mutually orthogonal Subtraction of equations (77) and (78) gives:

$$\frac{x^2}{(a^2+u)(a^2+v)} + \frac{y^2}{(b^2+u)(b^2+v)} + \frac{z^2}{(c^2+u)(c^2+v)} = 0. \tag{80}$$

The direction-cosines of the normals to surfaces (77) and (78) are respectively proportional to

$$\frac{x}{a^2+u}; \quad \frac{y}{b^2+u}; \quad \frac{z}{c^2+u}; \quad \text{and} \quad \frac{x}{a^2+v}; \quad \frac{y}{b^2+v}; \quad \frac{z}{c^2+v},$$

and equation (80) expresses the condition for these normals to be perpendicular, i.e. proves that surfaces (77) and (78) are orthogonal. The other two coordinate surfaces may similarly be shown to be orthogonal. Dupin's theorem enables us to state that *the two families of lines of curvature are obtained on the ellipsoid* (77) *(with fixed u) as the intersections of the ellipsoid with all the hyperboloids of families* (78) *and* (79).

138. Gaussian curvature. We shall explain the geometrical significance of the Gaussian curvature. We take the lines of curvature as coordinate lines on a surface. Relationship (72) will be satisfied along all these lines, the coefficient a being one of the principal radii of curvature, as we have seen. This gives us the following relationships:

$$\mathbf{r}'_u + R_1 \mathbf{m}'_u = 0; \quad \mathbf{r}'_v + R_2 \mathbf{m}'_v = 0. \tag{81}$$

With every point M of the surface we associate a point M_0 of the unit sphere, M_0 being the terminus of the unit vector $\mathbf{m}$ drawn from the centre of the sphere, where $\mathbf{m}$ is the unit normal to the

surface at M. This correspondence between points of a surface and points of a sphere is usually known as a *spherical mapping of the surface*. The position of the point M_0 will be determined by the same parameters u and v as determine M. Since the coordinate lines are the lines of curvature, we have:

$$E = \mathbf{r}_u'^2; \quad F = 0; \quad G = \mathbf{r}_v'^2. \tag{82}$$

The radius vector of the spherical mapping M_0 is, by definition, $\mathbf{m}$, and the coefficients of the first Gauss form for the spherical mapping are, by (81) and (82):

$$E_0 = \mathbf{m}_u'^2 = \frac{1}{R_1^2} E; \quad F_0 = \mathbf{m}_u' \cdot \mathbf{m}_v' = 0; \quad G_0 = \mathbf{m}_v'^2 = \frac{1}{R_2^2} G. \tag{83}$$

We shall only stop to prove the middle equation, since the others follow directly from (81) and (82). Expressions (49) give $M = = -\mathbf{r}_u' \cdot \mathbf{m}_v' = -\mathbf{r}_v' \cdot \mathbf{m}_u'$. Since we have taken the lines of curvature as coordinate lines, $M = 0$, i.e. $\mathbf{r}_u' \cdot \mathbf{m}_v' = \mathbf{r}_v' \cdot \mathbf{m}_u' = 0$. On multiplying the first of equations (81) by $\mathbf{m}_v'$ or the second by $\mathbf{m}_u'$, we get $\mathbf{m}_u' \cdot \mathbf{m}_v' = 0$.

An elementary area of the original surface, and the corresponding element of the spherical mapping, are given by

$$\mathrm{d}S = \sqrt{EG}\, \mathrm{d}u\, \mathrm{d}v; \quad \mathrm{d}S_0 = \sqrt{E_0 G_0}\, \mathrm{d}u\, \mathrm{d}v,$$

or, by (83),

$$\mathrm{d}S_0 = \frac{1}{|R_1 R_2|} \mathrm{d}S,$$

whence it is clear that *the Gaussian curvature at a point M has an absolute value equal to the limit of the ratio of an area of the spherical mapping to the corresponding area of the original surface when the latter contracts indefinitely to the point M*. This ratio obviously characterizes the degree of dispersion of the pencil of normals to the surface at points of the elementary area.

We obtained in [134] an expression for the Gaussian curvature K in terms of the coefficients of the two Gauss forms. The expression for K given by Gauss himself was only in terms of the coefficients E, F, G and their derivatives with respect to u and v. This fact has an important consequence, which we must stop to consider. Let there be a correspondence between points of two surfaces (S) and (S_1) such that corresponding points have the same values of parameters u, v. Each surface will have its own first Gauss form, expressing

the square of an element of arc. If the two forms are identical, it amounts to saying that lengths are preserved in the correspondence, or in other words, that *the surfaces can be superimposed on each other.* In this case, the coefficients E, F, G and their derivatives with respect to u and v are the same for both surfaces, so that the curvature K has the same value at corresponding points of the surfaces, i.e. *when a mapping of one surface onto another preserves lengths, the Gaussian curvature has the same value at corresponding points.*

In particular, the Gaussian curvature is zero on a plane, and we must have $LN - M^2 = 0$ on a surface that can be superimposed on a plane without distortion of lengths, i.e. every point must be parabolic. We have already had the cone and cylinder as examples of such surfaces.

139. The variation of an elementary area and the mean curvature. Let (u, v) be the parametric coordinates and $\mathbf{r}(u, v)$ the radius vector of a given surface (S). If we mark off along the normal $\mathbf{m}$ at every point $M(u, v)$ of the surface a length MM_1 of algebraic value $n(u, v)$, where $n(u, v)$ is a function of u and v, we get a new surface (S_1) formed by the points M_1. We shall represent points M_1 by the same parameters (u, v) as points M, and shall speak of a correspondence having been set up along the normals to (S) between points of (S) and (S_1). The radius vector $\mathbf{r}^{(1)}(u, v)$ to the surface (S_1) is by definition: $\mathbf{r}^{(1)}(u, v) = \mathbf{r}(u, v) + n(u, v)\,\mathbf{m}(u, v)$. We obtain on differentiating with respect to u and v:

$$\mathbf{r}_u^{(1)\prime} = \mathbf{r}_u' + n_u'\mathbf{m} + n\mathbf{m}_u'; \quad \mathbf{r}_v^{(1)\prime} = \mathbf{r}_v' + n_v'\mathbf{m} + n\mathbf{m}_v'.$$

We now find the coefficients E_1, F_1, G_1 of the first Gauss form for (S_1), on the assumption that the length n and its derivatives with respect to u and v are small so that second order terms may be neglected:

$$E_1 = (\mathbf{r}_u^{(1)\prime})^2 = (\mathbf{r}_u' + n_u'\mathbf{m} + n\mathbf{m}_u')\cdot(\mathbf{r}_u' + n_u'\mathbf{m} + n\mathbf{m}_u') =$$

$$= \mathbf{r}_u'^2 + 2n_u'(\mathbf{r}_u'\cdot\mathbf{m}) + 2n(\mathbf{r}_u'\cdot\mathbf{m}_u').$$

The vectors $\mathbf{r}_u'$ and $\mathbf{m}$ are perpendicular and $\mathbf{r}_u' \cdot \mathbf{m} = 0$, so that (47) gives $E_1 = E - 2nL$. Similarly, we find that $F_1 = F - 2nM$ and $G_1 = G - 2nN$. Hence:

$$E_1 G_1 - F_1^2 = EG - F^2 - 2n(EN - 2FM + GL),$$

or, by (67):

$$E_1 G_1 - F_1^2 = (EG - F^2)(1 - 4nH).$$

If we take square roots, expand $(1 - 4nH)^{1/2}$ by the binomial theorem and neglect higher powers of n than the first, we get:

$$\sqrt{E_1 G_1 - F_1^2} = \sqrt{EG - F^2}\,(1 - 2nH). \tag{84}$$

On multiplying by dudv and integrating, we get an expression for the difference δS between the areas of the neighbouring surfaces (S) and (S_1) to an accuracy of second order terms:

$$\iint\limits_{(S_1)} \sqrt{E_1 G_1 - F_1^2}\, \mathrm{d}u\, \mathrm{d}v - \iint\limits_{(S)} \sqrt{EG - F^2}\, \mathrm{d}u\, \mathrm{d}v =$$

$$= - \iint\limits_{(S)} 2nH \sqrt{EG - F^2}\, \mathrm{d}u \mathrm{d}v \tag{85}$$

or

$$\delta S = - \iint\limits_{(S)} 2nH\, \mathrm{d}S.$$

The familiar problem of Plateau, of *finding the surface of minimum area stretched on a given contour* (L), is directly connected with this expression. It is easily seen that *the mean curvature H must be zero on such a surface*. If H were say positive on some part σ of such a surface, we should obtain, by (85), on choosing a small n, also positive on σ and zero elsewhere, including in particular (L), a negative value for δS:

$$\delta S = - \iint\limits_{(\sigma)} 2nH\, \mathrm{d}S,$$

and the surface (S_1) passing through (L) would have an area less than (S), which contradicts our original hypothesis. In view of the above, a surface of zero mean curvature is known as a *minimal surface*.

The formula for differentiating an integral over a variable closed surface with respect to a parameter also follows from (84). Let the position of a variable closed surface be defined by a parameter λ, and let its position be (S) for $\lambda = \lambda_0$, and (S_1), near (S), for λ near λ_0. We set up a normal correspondence, as above, between points M of (S) and M_1 of (S_1). With this, n is a function of u, v, and λ, which vanishes as an identity in u and v for $\lambda = \lambda_0$, i.e.

$$n(u, v, \lambda_0) \equiv 0. \tag{86}$$

Further, let $f(N)$ be a function of points of space which is independent of the parameter λ. The value of the integral:

$$I(\lambda) = \iint_{(S_1)} f(M_1)\,\mathrm{d}S_1 \tag{87}$$

will depend on λ, since the form of the surface depends on the parameter. We find an expression for the derivative $I'(\lambda_0)$. On multiplying both sides of (84) by $\mathrm{d}u\mathrm{d}v$, we can write $\mathrm{d}S_1 = (1 - 2nH)\,\mathrm{d}S$, and (87) can be written as:

$$I(\lambda) = \iint_{(S)} f(M_1)\,\mathrm{d}S - \iint_{(S)} f(M_1)\,2nH\,\mathrm{d}S.$$

The domain of integration is now the original surface (S) and is no longer dependent on λ, and we can use the ordinary rule for differentiating under the integral sign [80]. Let the point M of (S) correspond to the point M_1 of (S_1), so that $\overline{MM_1} = n(u, v)$ is normal to (S), i.e. has the direction $\mathbf{m}$. Differentiation of $f(M_1)$ with respect to λ gives at $\lambda = \lambda_0$:

$$\lim_{\lambda\to\lambda_0} \frac{f(M_1) - f(M)}{\lambda - \lambda_0} = \lim_{\lambda\to\lambda_0} \frac{f(M_1) - f(M)}{\overline{MM_1}} \cdot \frac{\overline{MM_1}}{\lambda - \lambda_0} = \frac{\partial f(M)}{\partial m} \cdot \frac{\partial n}{\partial \lambda}\bigg|_{\lambda=\lambda_0}$$

where m is the direction of the normal $\mathbf{m}$. On noting that n vanishe for $\lambda = \lambda_0$ and writing $\partial n/\partial\lambda_0$ for the value of the derivative at $\lambda = \lambda_0$ we get:

$$I'(\lambda_0) = \iint_{(S)} \frac{\partial f(M)}{\partial m} \cdot \frac{\partial n}{\partial \lambda_0}\,\mathrm{d}S - \iint_{(S)} f(M)\,2H\frac{\partial n}{\partial\lambda_0}\,\mathrm{d}S. \tag{88}$$

Let the equation of the variable surface (S_1) be given implicitly as:

$$\varphi(M_1;\ \lambda) = 0 \quad \text{or} \quad \varphi(x, y, z, \lambda) = 0. \tag{89}$$

Differentiating with respect to λ both directly and via M_1, as in the case of the function $f(M_1)$, we get for $\lambda = \lambda_0$:

$$\frac{\partial\varphi(M_1\,\lambda_0)}{\partial\lambda_0} + \frac{\partial\varphi(M_1, \lambda_0)}{\partial m} \cdot \frac{\partial n}{\partial\lambda_0} = 0.$$

If we find $\partial n/\partial\lambda_0$ from this and substitute in (88), we get the following expression for the derivative:

$$I'(\lambda_0) = -\iint_{(S)} \frac{\partial f}{\partial m}\,\frac{\dfrac{\partial\varphi}{\partial\lambda_0}}{\dfrac{\partial\varphi}{\partial m}}\,\mathrm{d}S + 2\iint_{(S)} fH\,\frac{\dfrac{\partial\varphi}{\partial\lambda_0}}{\dfrac{\partial\varphi}{\partial m}}\,\mathrm{d}S. \tag{90}$$

If the integrand f in integral (87) also contains the parameter λ, an extra term has to be added on the right of (90), as was the case in [120], of the form:

$$\iint_{(S)} \frac{\partial f}{\partial \lambda_0} \, \mathrm{d}S.$$

140. Envelopes of surfaces and curves. We introduced the idea of an envelope of plane curves in [10], when studying the particular solutions of first order ordinary differential equations. In a similar way, the solution of partial differential equations leads us to the concept of an envelope of surfaces, a brief account of which now follows.

Let us be given a family of surfaces with a single parameter,

$$F(x, y, z, a) = 0. \tag{91}$$

A definite surface of the family is obtained on fixing the numerical value of a. We consider a new surface (S) which also has the equation (91), but with variable a, found from the equation:

$$\frac{\partial F(x, y, z, a)}{\partial a} = 0. \tag{92}$$

We can say that the equation of (S) is obtained by eliminating a from equations (91) and (92). If we take a fixed $a = a_0$, on the one hand we get a definite surface (S_0) of family (91), and on the other hand, the substitution of $a = a_0$ in (91) and (92) gives us a line (l_0) on the surface (S), such that (S) and (S_0) have (l_0) in common. We shall prove that the surfaces have a common tangent plane along (l_0).

Since a is constant, the projections $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ of an infinitesimal movement along the surface (91) must satisfy

$$\frac{\partial F}{\partial x} \mathrm{d}x + \frac{\partial F}{\partial y} \mathrm{d}y + \frac{\partial F}{\partial z} \mathrm{d}z = 0.$$

Since a is variable for surface (S), we must write here:

$$\frac{\partial F}{\partial x} \mathrm{d}x + \frac{\partial F}{\partial y} \mathrm{d}y + \frac{\partial F}{\partial z} \mathrm{d}z + \frac{\partial F}{\partial a} \mathrm{d}a = 0.$$

But this equation is the same as the previous one, by (92), i.e. infinitesimal displacements at a common point on (S_0) and (S) are perpendicular to the same direction, the direction-cosines of which are proportional to:

$$\frac{\partial F}{\partial x}, \quad \frac{\partial F}{\partial y}, \quad \frac{\partial F}{\partial z},$$

so that it follows that (S_0) and (S) touch along (l_0). Thus in general, *elimination of a from equations* (91) *and* (92) *gives the equation of the envelope of surfaces of family* (91), *with contact taking place along a line.*

Example. Let us take the family of spheres with centres on the z axis and radius r (constant):

$$x^2 + y^2 + (z - a)^2 = r^2.$$

We differentiate with respect to a:

$$-2(z - a) = 0.$$

On eliminating a, we get the equation of the circular cylinder:

$$x^2 + y^2 = r^2,$$

which touches each of the spheres in a circle.

We now consider a family of surfaces with two parameters:

$$F(x, y, z, a, b) = 0. \tag{93}$$

On eliminating a and b from this equation and the equations

$$\frac{\partial F(x, y, z, a, b)}{\partial a} = 0; \qquad \frac{\partial F(x, y, z, a, b)}{\partial b} = 0, \tag{94}$$

it is easily shown that we get a surface (S), which touches the surfaces of family (93). But in this case, contact only takes place at a point, instead of along a line. In fact, on taking a fixed $a = a_0$ and $b = b_0$, on the one hand we get a definite surface (S_0) of family (93), and on the other hand, substitution of $a = a_0$ and $b = b_0$ in the three equations (93) and (94) in general gives a point M_0 on the surface (S). The point M_0 will be common to (S) and (S_0).

Example. Let us take the family of spheres with centres on the XY plane and a fixed radius r:

$$(x - a)^2 + (y - b)^2 + z^2 = r^2.$$

We differentiate with respect to a and b:

$$-2(x - a) = 0; \quad -2(y - b) = 0;$$

Elimination of a and b gives us the equation $z^2 = r^2$, i.e. the envelope consists of two parallel planes $z = \pm r$, which touch each of the spheres at a point.

The remark regarding the determination of the envelope of a family of curves [10] applies equally when finding the envelope of a family of surfaces: elimination of a from equations (91) and (92),

for instance, can lead not only to the envelope but also to the locus of singular points of the surfaces of family (91), i.e. those points at which the surfaces have no tangent plane. If the left-hand side of (91) is continuous and has continuous first order derivatives, every surface that touches the various surfaces of family (91) at all its points can be found by the above method of eliminating a from equations (91) and (92). In general, we omit proofs and precise conditions in this article and the next, and confine ourselves to giving the broad outlines of basic facts.

We now consider a family of curves in space, depending on a single parameter:

$$F_1(x, y, z, a) = 0; \qquad F_2(x, y, z, a) = 0. \tag{95}$$

We shall seek the envelope of the family, i.e. the curve Γ, every point of which is a point of contact with a curve of family (95). We can take Γ as also defined by equations (95), except that now a is variable instead of constant [10]. The projections $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ of a small displacement along curves (95) must satisfy the equations

$$\frac{\partial F_1}{\partial x}\,\mathrm{d}x + \frac{\partial F_1}{\partial y}\,\mathrm{d}y + \frac{\partial F_1}{\partial z}\,\mathrm{d}z = 0;$$
$$\frac{\partial F_2}{\partial x}\,\mathrm{d}x + \frac{\partial F_2}{\partial y}\,\mathrm{d}y + \frac{\partial F_2}{\partial z}\,\mathrm{d}z = 0.$$

Similarly, the projections $\delta x, \delta y, \delta z$ of a small displacement along Γ must satisfy the equations:

$$\frac{\partial F_1}{\partial x}\,\delta x + \frac{\partial F_1}{\partial y}\,\delta y + \frac{\partial F_1}{\partial z}\,\delta z + \frac{\partial F_1}{\partial a}\,\delta a = 0$$
$$\frac{\partial F_2}{\partial x}\,\delta x + \frac{\partial F_2}{\partial y}\,\delta y + \frac{\partial F_2}{\partial z}\,\delta z + \frac{\partial F_2}{\partial a}\,\delta a = 0.$$

The condition for contact amounts to these projections being proportional, i.e.

$$\frac{\delta x}{\mathrm{d}x} = \frac{\delta y}{\mathrm{d}y} = \frac{\delta z}{\mathrm{d}z},$$

and this in turn, in view of the above relationships, is equivalent to the two equations: $(\partial F_1/\partial a)\,\delta a = 0$ and $(\partial F_2/\partial a)\,\delta a = 0$, or, if we take $\delta a \neq 0$, i.e. a not a constant, we get the two equations:

$$\frac{\partial F_1(x, y, z, a)}{\partial a} = 0; \qquad \frac{\partial F_2(x, y, z, a)}{\partial a} = 0. \tag{96}$$

The four equations (95) and (96) do not in general define a curve, i.e. as a rule, *a family of curves in space has no envelope*. But if the

four equations reduce to three, i.e. one is a consequence of the others, the coordinates (x, y, z) will be defined by the three equations as functions of the parameter a, so that we get a curve in space, which is in fact an envelope [or the locus of singular points of curves (95)]. We have an example in the next article of a family of straight lines in space which has an envelope.

141. Developable surfaces. We take the particular case of a family of planes with one parameter a:

$$A(a)\,x + B(a)\,y + C(a)\,z + D(a) = 0, \tag{97}$$

The envelope (S) is obtained by eliminating a from the two equations:

$$\left.\begin{aligned} A(a)\,x + B(a)\,y + C(a)z + D(a) &= 0 \\ A'(a)\,x + B'(a)\,y + C'(a)z + D'(a) &= 0. \end{aligned}\right\} \tag{98}$$

With a fixed, these equations yield a straight line (l_a), and the surface (S) is the locus of these straight lines, i.e. (S) must be a ruled surface. It may further be seen that not every ruled surface can be obtained by the above method. The surface (S) touches a plane (97) along (l_a), i.e. (S) *has the same tangent plane along the straight generator* (l_a). Thus a family of tangent planes on (S) depends only on the single parameter a, designating the generator (l_a). The family of tangent planes to a surface depends in the general case on two parameters, defining the position of a point on the surface. Let the equation of (S) be written explicitly: $z = f(x, y)$, the partial derivatives of the function $f(x, y)$ being denoted as in [62]. The first two direction-cosines of the normal will be functions of the single parameter a:

$$\frac{p}{\sqrt{1+p^2+q^2}} = W_1(a); \quad \frac{q}{\sqrt{1+p^2+q^2}} = W_2(a).$$

Elimination of a from these equations gives a relationship between p and q that can be written as:

$$q = \varphi(p).$$

This relationship must be satisfied over the entire surface (S) and we find on differentiating with respect to the independent variables x and y:

$$s = \varphi'(p)\,r; \quad t = \varphi'(p)\,s,$$

whence

$$rt - s^2 = 0, \tag{99}$$

i.e. *all the points on a surface enveloping a family of planes with one parameter must be parabolic.*

The surface (S) is generated by the family of straight lines (98). It is easily seen that this family has an envelope; differentiation of equations (98) with respect to a gives the two equations:

$$\left.\begin{aligned} A'(a)\,x + B'(a)\,y + C'(a)\,z + D'(a) = 0, \\ A''(a)\,x + B''(a)\,y + C''(a)\,z + D''(a) = 0, \end{aligned}\right\} \tag{100}$$

and the four equations (98) and (100) reduce to three. We can therefore say that (S) is generated by tangents to the curve Γ in space. If the curve Γ degenerates to a point, (S) is a conical surface, whilst it is cylindrical if the point is at infinity. We prove the converse: given a curve Γ in space,

$$x = \varphi(t); \quad y = \psi(t); \quad z = \omega(t), \tag{101}$$

the surface (S) generated by tangents to Γ envelopes a family of planes with one parameter, these being the osculating planes of Γ. The family has in fact the equation

$$A(X - x) + B(Y - y) + C(Z - z) = 0, \tag{102}$$

where (x, y, z) are given by equations (101) and A, B, C are given by expressions (31) of [126]. On differentiating (102) with respect to the parameter t and using the fact that, by (31),

$$A\,\mathrm{d}x + B\,\mathrm{d}y + C\,\mathrm{d}z = 0, \tag{103}$$

we get

$$\mathrm{d}A\,(X - x) + \mathrm{d}B\,(Y - y) + \mathrm{d}C(Z - z) = 0, \tag{104}$$

where we write the differentials instead of the derivatives with respect to t. The enveloping surface of family (102) is made up of straight lines determined by equations (102) and (104), and it remains for us to show that these two equations give the tangent to Γ at the point (x, y, z). We differentiate (103) and note that $A\mathrm{d}^2x + B\mathrm{d}^2y + C\mathrm{d}^2z = 0$ by (31); we get:

$$\mathrm{d}A\,\mathrm{d}x + \mathrm{d}B\,\mathrm{d}y + \mathrm{d}C\,\mathrm{d}z = 0. \tag{105}$$

Equations (103) and (105) show that normals to the planes (102) and (104), which pass through the point (x, y, z), are perpendicular to the tangent to the curve Γ, i.e. planes (102) and (104) both pass through this tangent, which is what we wished to prove.

We saw above that condition (99) is necessary for (S) to be the envelope of a family of planes with one parameter. It may be shown to be also sufficient. We also mentioned above [138] that (99) (or its equivalent $LN - M^2 = 0$) is necessary for (S) to be able to be mapped on to a plane without distortion of length. The converse can be proved, that if this condition is fulfilled, a sufficiently small portion of the surface can be mapped on to a plane by the method described. For this reason, the envelope of a family of planes with one parameter is known as a *developable surface*.

Not every ruled surface is developable. For instance, if we take a hyperbolic paraboloid or a hyperboloid of one sheet, (99) is not fulfilled for these [137], in spite of the fact that they are ruled surfaces. It follows from this that, if a point varies along a straight generator of such a surface, the corresponding tangent plane rotates about the generator.

The French mathematician Lebesgue carried out a detailed investigation of surfaces developable into planes, with very few assumptions regarding the functions appearing in equations (38) of the surface (we have assumed the existence of continuous derivatives up to the second order). One of his results was a developable surface consisting of a non-ruled surface of revolution.

CHAPTER VI

FOURIER SERIES

§ 14. Harmonic analysis

142. Orthogonality of the trigonometric functions. The harmonic oscillation

$$y = A \sin(\omega t + \varphi)$$

represents the simplest example of a periodic function of period $T = 2\pi/\omega$. We confine ourselves for the present to periodic functions of period 2π and let x denote the independent variable, so that the function y becomes:

$$y = A \sin(x + \varphi).$$

More complicated functions of the same period are given by

$$A_k \sin(kx + \varphi_k) \quad (k = 0, 1, 2, 3, \dots),$$

or by the sum of any number of these:

$$\sum_{k=0}^{n} A_k \sin(kx + \varphi_k),$$

this sum being known as a *trigonometric polynomial of the n-th order*. The question now naturally arises of *the approximate representation of any periodic function $f(x)$ of period 2π as a trigonometric polynomial of the n-th order*, followed by the question of *the expansion of $f(x)$ into a trigonometric series:*

$$f(x) = \sum_{k=0}^{\infty} A_k \sin(kx + \varphi_k);$$

these problems are similar to the problems of approximating a function by a polynomial of the nth degree or of expanding it into a power series. The general term of the above series,

$$A_k \sin(kx + \varphi_k),$$

is known as the *k-th harmonic of the function* $f(x)$. It can be written in the form

$$A_k \sin(kx + \varphi_k) = a_k \cos kx + b_k \sin kx,$$

where

$$a_k = A_k \sin \varphi_k; \quad b_k = A_k \cos \varphi_k \quad (k = 0, 1, 2, \dots).$$

The zero order harmonic, $A_0 \sin \varphi_0$, is simply a constant, which we denote by $a_0/2$ in order to simplify later expressions. Our problem thus amounts to *choosing, if possible, the unknown constants*

$$a_0, a_1, b_1, a_2, b_2 \dots, a_n, b_n, \dots$$

in such a way that the series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \tag{1}$$

is convergent to a sum equal to the given periodic function $f(x)$ *of period* 2π.

As a preliminary to solving our problem, we note a simple property of the sines and cosines of multiple angles. Let c be any real number, and $(c, c + 2\pi)$ an interval of length 2π. It is easy to show that

$$\int_c^{c+2\pi} \cos kx \,\mathrm{d}x = 0; \qquad \int_c^{c+2\pi} \sin kx \,\mathrm{d}x = 0 \quad (k = 1, 2, 3, \dots). \tag{2}$$

If we take say the first of the integrals written, the primitive for $\cos kx$ is $(1/k) \sin kx$, and in view of its periodicity, its values are the same for $x = c$ and $x = c + 2\pi$, so that their difference is zero, i.e. in fact,

$$\int_c^{c+2\pi} \cos kx \,\mathrm{d}x = \frac{\sin kx}{k} \Big|_{x=c}^{x=c+2\pi} = 0.$$

Similary, by using the familiar trigonometric formulae:

$$\sin kx \cos lx = \frac{\sin(k + l)\, x + \sin(k - l)\, x}{2},$$

$$\sin kx \sin lx = \frac{\cos(k - l)\, x - \cos(k + l)\, x}{2},$$

$$\cos kx \cos lx = \frac{\cos(k + l)\, x + \cos(k - l)\, x}{2},$$

it can be shown that:

$$\int_c^{c+2\pi} \cos kx \sin lx \,\mathrm{d}x = 0; \qquad \int_c^{c+2\pi} \cos kx \cos lx \,\mathrm{d}x = 0;$$

$$\int_c^{c+2\pi} \sin kx \sin lx \,\mathrm{d}x = 0 \qquad (k \neq l). \tag{3}$$

Let us take the family of functions

$$1, \cos x, \ \sin x, \ \cos 2x, \ \sin 2x, \ldots, \ \cos nx, \ \sin nx, \ldots, \tag{4}$$

the first member of the family being a constant equal to unity. Formulae (2) and (3) express the following fact: *the integral of the product of any two different functions of family* (4) *over any interval of length* 2π *is zero.* This property is generally known as the *orthogonality of family* (4) *in the interval mentioned.* We now find the integral of the square of a function of the family. The integral is evidently equal to 2π for the first function, whilst for the remainder, since

$$\cos^2 kx = \frac{1 + \cos 2kx}{2}\ ; \quad \sin^2 kx = \frac{1 - \cos 2kx}{2}\ ,$$

we have:

$$\int\limits_c^{c+2\pi} \cos^2 kx \, dx = \pi; \quad \int\limits_c^{c+2\pi} \sin^2 kx \, dx = \pi \quad (k = 1, \ 2, \ldots). \tag{5}$$

To avoid confusion, we shall in future take $c = -\pi$, i.e. the interval $(c, c + 2\pi)$ now becomes $(-\pi, \pi)$.

We now turn to the problem posed above. Let the function $f(x)$ be defined in the interval $(-\pi, \pi)$, in which case it is defined for other values of x by virtue of its periodicity of period 2π; and let us assume that it gives the sum of series (1):

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx). \tag{6}$$

On integrating both sides of this equation over the interval $(-\pi, \pi)$ and replacing the integral of the infinite sum by the sum of the separate integrals, we get:

$$\int\limits_{-\pi}^{+\pi} f(x) \, dx = \int\limits_{-\pi}^{+\pi} \frac{a_0}{2} \, dx + \sum_{k=1}^{\infty} \left(a_k \int\limits_{-\pi}^{+\pi} \cos kx \, dx + b_k \int\limits_{-\pi}^{+\pi} \sin kx \, dx \right),$$

which, by (2), reduces to the equation:

$$\int\limits_{-\pi}^{+\pi} f(x) \, dx = \frac{a_0}{2} \cdot 2\pi = a_0 \pi,$$

whence we determine the constant a_0:

$$a_0 = \frac{1}{\pi} \int\limits_{-\pi}^{+\pi} f(x) \, dx. \tag{7}$$

We now have to find the remaining constants. Let n be a positive integer; let us multiply both sides of (6) by $\cos nx$ and integrate, as above:

$$\int_{-\pi}^{+\pi} f(x)\cos nx\,\mathrm{d}x = \frac{a_0}{2}\int_{-\pi}^{+\pi}\cos nx\,\mathrm{d}x +$$

$$+\sum_{k=1}^{\infty}\left(a_k\int_{-\pi}^{+\pi}\cos kx\cos nx\,\mathrm{d}x + b_k\int_{-\pi}^{+\pi}\sin kx\cos nx\,\mathrm{d}x\right). \tag{8}$$

All the integrals on the right of the equation are zero by (2) and (3), with the exception of the one integral:

$$\int_{-\pi}^{+\pi}\cos kx\cos nx\,\mathrm{d}x \text{ with } k = n,$$

this latter integral being equal to π by (5).

Equation (8) thus reduces to the form:

$$\int_{-\pi}^{+\pi} f(x)\cos nx\,\mathrm{d}x = a_n\pi,$$

whence

$$a_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\cos nx\,\mathrm{d}x \qquad (n = 1, 2, \ldots). \tag{7_1}$$

In exactly the same way, we can obtain the formula:

$$b_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\sin nx\,\mathrm{d}x \quad (n = 1, 2, \ldots). \tag{7_2}$$

It may be noted that (7) is the same as (7_1) with $n = 0$. We can thus write:

$$\left.\begin{aligned} a_k &= \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\cos kx\,\mathrm{d}x \quad (k = 0, 1, 2, \ldots) \\ b_k &= \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\sin kx\,\mathrm{d}x \quad (k = 1, 2, \ldots). \end{aligned}\right\} \tag{9}$$

The above working is not rigorous and is only valuable as a guide. We have in fact made a number of unjustified assumptions: firstly, we assumed right at the start that the given function had the expansion (6), then we replaced the integral of the infinite sum by the sum

of the integrals of the separate terms, or as we say, integrated term by term, which is not always permissible [cf. I, 146].

The rigorous statement of the problem is as follows. Let a function $f(x)$ be given in the interval $(-\pi, \pi)$. We evaluate constants a_k and b_k in accordance with (9) and substitute the values obtained in series (1). The question arises: will the series thus obtained be convergent in the interval $(-\pi, \pi)$, and if so, will its sum be equal to $f(x)$?

The coefficients a_k and b_k obtained from (9) are known as the *Fourier coefficients of the function* $f(x)$, whilst the series obtained from (1) after replacing the a_k and b_k with their values as given by (9) is called the *Fourier series for* $f(x)$. We state in the next article the solution of the above problem of the convergence of the Fourier series for a given function.

Remark. Expressions (3) and (5) given above are valid for integration over any interval of length 2π. In general, if a function $f(x)$, defined for all real values of x, has a period a, i.e. $f(x + a) = f(x)$ for all x, the integral of $f(x)$ over any interval of length a has a definite value, independent of the initial point of the interval, i.e. the value of

$$\int_c^{c+a} f(x)\,\mathrm{d}x$$

is independent of c. The number c can in fact be written in the form $c = ma + h$, where m is an integer and h belongs to the interval $(0, a)$:

$$\int_c^{c+a} f(x)\,\mathrm{d}x = \int_{ma+h}^{(m+1)a+h} f(x)\,\mathrm{d}x = \int_{ma+h}^{(m+1)a} f(x)\mathrm{d}x + \int_{(m+1)a}^{(m+1)a+h} f(x)\,\mathrm{d}x.$$

We introduce a new variable of integration $t_1 = x - ma$ into the first integral, and $t_2 = x - (m + 1)\,a$ into the second:

$$\int_c^{c+a} f(x)\,\mathrm{d}x = \int_h^a f(t_1 + ma)\,\mathrm{d}t_1 + \int_0^h f[t_2 - (m + 1)\,a]\,\mathrm{d}t_2.$$

On taking account of the periodicity of $f(x)$ and denoting the variables of integration again by x, we get:

$$\int_c^{c+a} f(x)\,\mathrm{d}x = \int_h^a f(x)\,\mathrm{d}x + \int_0^h f(x)\,\mathrm{d}x = \int_0^a f(x)\,\mathrm{d}x,$$

whence it follows that the integral does not depend on c. If $f(x)$ has period 2π, we can evaluate its Fourier coefficients a_k and b_k in accordance with (9) by integrating over any interval of length 2π.

143. Dirichlet's theorem. The Fourier series for a function $f(x)$ will be convergent and its sum will be equal to $f(x)$ provided certain restrictions are imposed on $f(x)$. We suppose firstly that $f(x)$, given in the interval $(-\pi, \pi)$, is either continuous or has only a finite number of points of discontinuity in the interval. We further assume that all these discontinuities have the following property: if $x = c$ is a point of discontinuity of $f(x)$, there exist finite limits for $f(x)$ as x tends to c, both from the right (from larger values) and from the left (from smaller values). These limits are usually written as $f(c + 0)$ and $f(c - 0)$ [I, 32]. Such points of discontinuity are generally known as *discontinuities of the first kind.* We finally assume that the total interval $(-\pi, \pi)$ can be divided into a finite number of parts such that $f(x)$ varies monotonically in each. The above are generally referred to as *Dirichlet conditions*, i.e. we say that *a function satisfies Dirichlet conditions in the interval $(-\pi, \pi)$ if it is either continuous in the interval or has a finite number of discontinuities of the first kind, and if, furthermore, the interval can be divided into a finite number of sub-intervals in each of which $f(x)$ varies monotonically.* At the end $x = -\pi$, we are only interested in the limit to which $f(x)$ tends as x tends to $(-\pi)$ from the right, so that we shall write $f(-\pi + 0)$ instead of $f(-\pi)$; and similarly, instead of $f(\pi)$ we write $f(\pi - 0)$. We remark that these limits can be different, but the sum of series (1) must of course be the same for $x = -\pi$ and $x = \pi$, due to the periodicity of functions (4).

The following theorem is fundamental as regards the theory of Fourier series:

DIRICHLET'S THEOREM. *If $f(x)$ is specified and satisfies Dirichlet conditions in the interval $(-\pi, \pi)$, the Fourier series for the function is convergent throughout the interval and the sum of the series:*

(1) *is equal to $f(x)$ at all points of continuity of $f(x)$ lying in the interval;*

(2) *is equal to*

$$\frac{f(x+0)+f(x-0)}{2}$$

at all points of discontinuity;

(3) *is equal to*

$$\frac{f(-\pi+0)+f(\pi-0)}{2}$$

at the ends of the interval, i.e. *for $x = -\pi$ and $x = +\pi$.*

The proof of this theorem will be given at the end of the present chapter.

Certain points may be noted in regard to the statement of the theorem. The terms of series (1) are periodic functions with period 2π. Hence, if the series is convergent in the interval $(-\pi, \pi)$, it is likewise convergent for all real values of x, and the sum of the series periodically repeats, with period 2π, the values that it gave in $(-\pi, \pi)$. We must therefore assume, if we wish to use the Fourier series outside the interval $(-\pi, \pi)$, that the function $f(x)$ is continued outside with a periodicity of period 2π. The ends of the interval $x = \pm\pi$ are from this point of view points of discontinuity of the continued function, if $f(-\pi + 0) \neq f(\pi - 0)$.

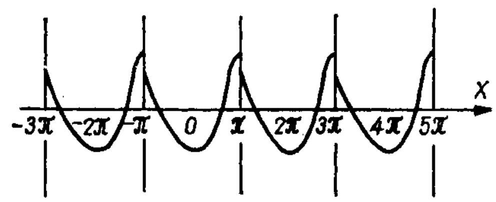

FIG. 114

A function continuous in $(-\pi, \pi)$ is illustrated in Fig. 114, which gives discontinuities on periodic continuation due to having different values at the ends of the interval.

The following lemma is often useful when calculating Fourier coefficients:

LEMMA. *If $f(x)$ is an even function in the interval* $(-a, a)$, i.e. $f(-x) = f(x)$, *we have*

$$\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx,$$

whilst if $f(x)$ is an odd function, i.e. $f(-x) = -f(x)$, *we have*

$$\int_{-a}^{a} f(x)\,dx = 0.$$

The proof of this lemma was given earlier [I, 99].

144. Examples. 1. We expand x as a Fourier series in the interval $(-\pi, \pi)$. The products $x \cos kx$ are odd functions of x, so that all the coefficients a_k are zero by (9). On the other hand, the products $x \sin kx$ are even functions, and the coefficients b_k may be evaluated from the formula:

$$b_k = \frac{2}{\pi}\int_{0}^{\pi} x \sin kx\,dx = \frac{2}{\pi}\left\{-\frac{x\cos kx}{k}\Bigg|_{x=0}^{x=\pi} + \frac{1}{k}\int_{0}^{\pi}\cos kx\,dx\right\} = \frac{2(-1)^{k-1}}{k}.$$

The graph of the Fourier series is drawn with a full line in Fig. 115, and it is clear from the figure that we have discontinuities at $x = \pm\pi$, the arithmetic mean of the limits from the left and right being evidently zero. Dirichlet's theorem thus gives in the present case:

$$2\left(\frac{\sin x}{1} - \frac{\sin 2x}{2} + \ldots + \frac{(-1)^{k-1}\sin kx}{k} + \ldots\right) = \begin{cases} x \text{ for } -\pi < x < \pi. \\ 0 \text{ for } x = \pm\pi. \end{cases} \tag{10}$$

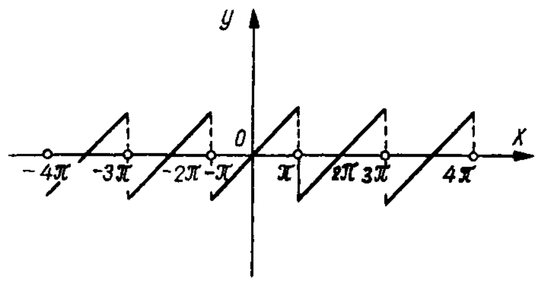

Fig. 115

2. We do the same for x^2. Here, the $x^2 \sin kx$ are odd functions and all the b_k are zero. We calculate the a_k:

$$a_0 = \frac{2}{\pi}\int_0^\pi x^2\,dx = \frac{2}{\pi}\frac{x^3}{3}\Bigg|_{x=0}^{x=\pi} = \frac{2\pi^2}{3};$$

$$a_k = \frac{2}{\pi}\int_0^\pi x^2 \cos kx\,dx = \frac{2}{\pi}\left\{\frac{x^2 \sin kx}{k}\Bigg|_{x=0}^{x=\pi} - \frac{2}{k}\int_0^\pi x \sin kx\,dx\right\} =$$

$$= \frac{4}{\pi k}\left\{\frac{x\cos kx}{k}\Bigg|_{x=0}^{x=\pi} - \frac{1}{k}\int_0^\pi \cos kx\,dx\right\} = (-1)^k\frac{4}{k^2}.$$

It is clear from Fig. 116 that the graph of the Fourier series has no discontinuities, and the sum of the series is equal to x^2 throughout the interval, including its ends:

$$x^2 = \frac{\pi^2}{3} + 4\sum_{k=1}^{\infty}(-1)^k\frac{\cos kx}{k^2} \qquad (-\pi \leqslant x \leqslant \pi). \tag{11}$$

On setting $x = 0$, we get:

$$1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \ldots + (-1)^{k-1}\frac{1}{k^2} + \ldots = \frac{\pi^2}{12}. \tag{12}$$

If we put

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \ldots = \sigma,$$

$$1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \ldots = \sigma_1, \tag{13}$$

we evidently have

$$\sigma = \sigma_1 + \frac{1}{4} + \frac{1}{16} + \frac{1}{36} + \ldots = \sigma_1 + \frac{1}{4}\sigma, \quad \sigma_1 = \frac{3}{4}\sigma,$$

and equation (12) gives

$$1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \ldots = \sigma_1 - \frac{1}{4}\sigma = \frac{1}{2}\sigma = \frac{\pi^2}{12},$$

i.e.

$$\sigma = 1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2} + \ldots = \frac{\pi^2}{6}, \tag{14}$$

$$\sigma_1 = 1 + \frac{1}{9} + \frac{1}{35} + \ldots + \frac{1}{(2n+1)^2} + \ldots = \frac{\pi^2}{8}.$$

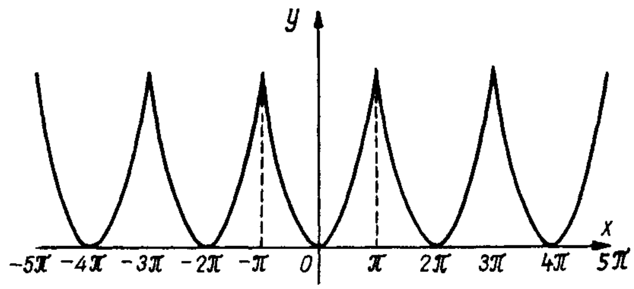

FIG. 116

3. We expand as a Fourier series the function:

$$f(x) = \begin{cases} c_1 & \text{for } -\pi < x < 0 \\ c_2 & \text{for } \quad 0 < x < \pi. \end{cases}$$

We have here

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\,dx = \frac{1}{\pi}\left[\int_{-\pi}^{0} c_1\,dx + \int_{0}^{\pi} c_2\,dx\right] = c_1 + c_2,$$

$$a_k = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\cos kx\,dx = \frac{1}{\pi}\left[\int_{-\pi}^{0} c_1 \cos kx\,dx + \right.$$

$$\left. + \int_{0}^{+\pi} c_2 \cos kx\,dx\right] = 0,$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\sin kx\,dx = \frac{1}{\pi}\left[\int_{-\pi}^{0} c_1 \sin kx\,dx + \int_{0}^{\pi} c_2 \sin kx\,dx\right] =$$

$$= (c_1 - c_2)\,\frac{(-1)^k - 1}{\pi k},$$

i.e. $b_k = 0$ for even k and $b_k = -2(c_1 - c_2)/\pi k$ for odd k, so that by Dirichlet's theorem (Fig. 117):

$$\frac{c_1 + c_2}{2} - \frac{2(c_1 - c_2)}{\pi}\left[\frac{\sin x}{1} + \frac{\sin 3x}{3} + \cdots\right] = \begin{cases} c_1 & \text{for} \quad -\pi < x < 0 \\ c_2 & \text{for} \quad 0 < x < \pi \\ \dfrac{c_1 + c_2}{2} & \text{for } x = 0 \text{ and } \pm\pi. \end{cases} \tag{15}$$

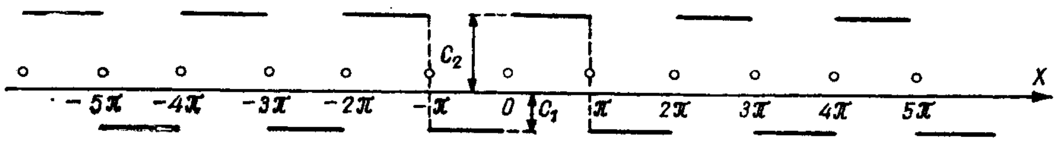

FIG. 117

145. Expansion in the interval $(0, \pi)$. We have simplified the evaluation of the Fourier coefficients in the above examples by making use of the evenness or oddness of the expanded $f(x)$.

In general, on applying the lemma of [143] to integrals (9) defining the Fourier coefficients, we get:

$$a_k = \frac{2}{\pi}\int_0^{\pi} f(x)\cos kx\,dx; \quad b_k = 0, \tag{16}$$

if $f(x)$ is an even function, and

$$a_k = 0; \quad b_k = \frac{2}{\pi}\int_0^{\pi} f(x)\sin kx\,dx, \tag{17}$$

if $f(x)$ is odd. The actual expansion of the function will be of the form

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx, \tag{18}$$

if $f(x)$ is even, and

$$\sum_{k=1}^{\infty} b_k \sin kx, \tag{19}$$

if $f(x)$ is odd.

Now let any function $f(x)$ be given in the interval $(0, \pi)$. It can be expanded in $(0, \pi)$ either in a series of the form (18) containing only cosines, or in a series of form (19) containing only sines. The coefficients are evaluated in accordance with formulae (16) in the first case, and in accordance with (17) in the second case. Both series have a sum inside the interval equal to $f(x)$, or to the arithmetic

mean at points of discontinuity. Outside $(0, \pi)$, however, they represent quite different functions: the cosine series gives a function obtained from $f(x)$ by even continuation in the neighbouring interval $(-\pi, 0)$, followed by periodic continuation with period 2π outside the interval $(-\pi, \pi)$. The sine series gives the function obtained by odd continuation

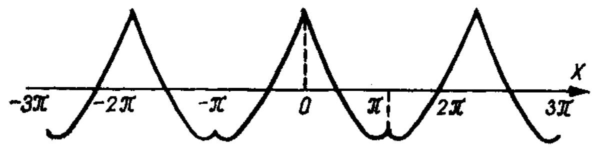

FIG. 118

in the neighbouring $(-\pi, 0)$, followed by periodic continuation with period 2π outside $(-\pi, \pi)$.

x	series in cos	series in sin
0	$f(+0)$	0
π	$f(\pi - 0)$	0

Thus in the cosine expansion:

$$f(-0) = f(+0);$$
$$f(-\pi + 0) = f(\pi - 0),$$

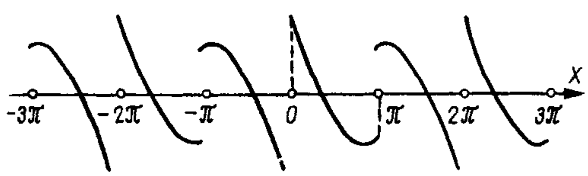

FIG. 119

whilst in the sine expansion:

$$f(-0) = -f(+0);$$
$$f(-\pi + 0) = -f(\pi - 0).$$

Correspondingly, we obtain at the ends of the interval the values shown in the table for series (18) and (19).

Figures 118 and 119 illustrate the graphs of the functions represented by series (18) and (19), derived from the same function $f(x)$ in the interval $(0, \pi)$.

Examples. **1.** We obtained in examples 1 and 2 of [144] a sine series for the function x and a cosine series for x^2 in $(0, \pi)$. On expanding x as a cosine series in $(0, \pi)$, we get:

$$x = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k \cos kx; \quad a_0 = \frac{2}{\pi}\int_0^{\pi} x\,dx = \pi;$$

$$a_k = \frac{2}{\pi}\int_0^{\pi} x\cos kx\,dx = \frac{2}{\pi k^2}\left[(-1)^k - 1\right] = \begin{cases} 0 & \text{for even } k \\ -\dfrac{4}{\pi k^2} & \text{for odd } k. \end{cases}$$

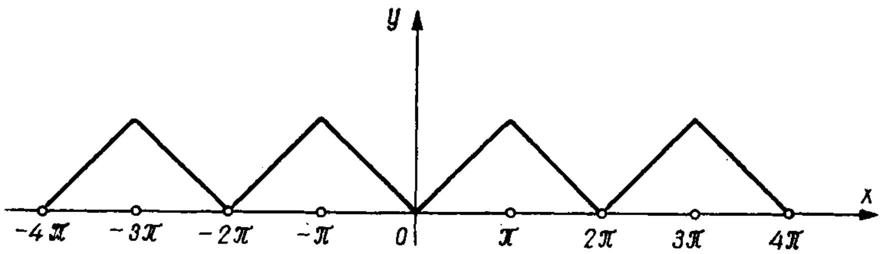

FIG. 120

Hence:

$$x = \frac{\pi}{2} - \frac{4}{\pi}\left(\frac{\cos x}{1^2} + \frac{\cos 3x}{3^2} + \ldots + \frac{\cos(2k+1)x}{(2k+1)^2} + \ldots\right) \tag{20}$$

$$(0 < x < \pi).$$

The sum of the series on the right will amount to $(-x)$ in the interval $(-\pi, 0)$, i.e. the sum gives the absolute value $|x|$ throughout the interval $(-\pi, \pi)$:

$$|x| = \frac{\pi}{2} - \frac{4}{\pi}\left(\frac{\cos x}{1^2} + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \ldots\right), \tag{21}$$

whilst outside this interval it gives the function obtained by periodic repetition of $|x|$ from $(-\pi, \pi)$ (Fig. 120). The sine expansion of x^2 in $(0, \pi)$ gives:

$$b_k = \frac{2}{\pi}\int_0^{\pi} x^2 \sin kx\,dx = \frac{2(-1)^{k-1}\pi}{k} + \frac{4\left[(-1)^k - 1\right]}{\pi k^3}$$

and

$$x^2 = 2\pi\left[\frac{\sin x}{1} - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \ldots\right] - \frac{8}{\pi}\left[\frac{\sin x}{1^3} + \frac{\sin 3x}{3^3} + \right.$$
$$\left. + \frac{\sin 5x}{5^3} + \ldots\right]$$

in the interval $0 < x < \pi$ (Fig. 121).

We suggest that the reader prove that we can split the Fourier series obtained into two series, as has been done above.

2. The function $\cos zx$ is even in regard to x, so that it can be expanded in cosines in the interval $(-\pi, \pi)$:

$$\cos zx = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx; \quad a_k = \frac{2}{\pi}\int_0^{\pi} \cos zx \cos kx \, dx.$$

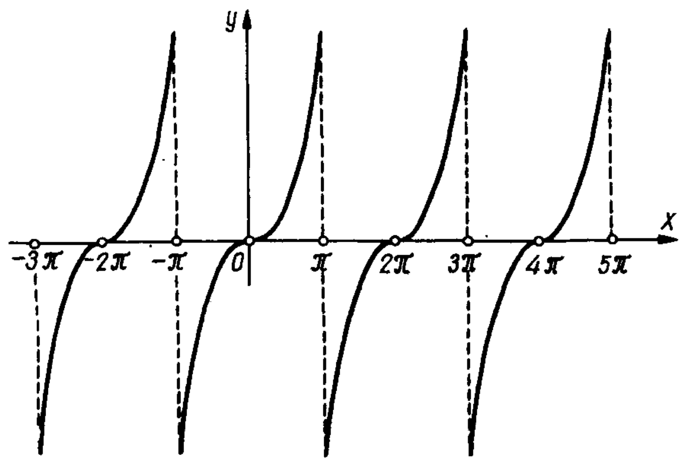

FIG. 121

We have:

$$a_0 = \frac{2}{\pi}\int_0^{\pi} \cos zx \, dx = \frac{2}{\pi}\,\frac{\sin zx}{z}\Big|_{x=0}^{x=\pi} = \frac{2 \sin \pi z}{\pi z};$$

$$a_k = \frac{2}{\pi}\int_0^{\pi} \cos zx \cos kx \, dx = \frac{1}{\pi}\int_0^{\pi} \left[\cos (z+k)\,x + \cos (z-k)\,x\right] dx =$$

$$= \frac{1}{\pi}\left[\frac{\sin (z+k)\,x}{z+k} + \frac{\sin (z-k)\,x}{z-k}\right]_{x=0}^{x=\pi} =$$

$$= \frac{1}{\pi}\left[\frac{\sin (\pi z + k\pi)}{z+k} + \frac{\sin (\pi z - k\pi)}{z-k}\right] = (-1)^k \frac{2z \sin \pi z}{\pi (z^2 - k^2)}.$$

Accordingly, in the interval $-\pi < x < \pi$:

$$\cos zx = \frac{2z \sin \pi z}{\pi}\left[\frac{1}{2z^2} + \frac{\cos x}{1^2 - z^2} - \frac{\cos 2x}{2^2 - z^2} + \frac{\cos 3x}{3^2 - z^2} - \ldots\right].$$

On setting $x = \pi$ here and dividing both sides of the equation by $\sin \pi z$, we have:

$$\cot \pi z = \frac{1}{\pi}\left[\frac{1}{z} - \sum_{k=1}^{\infty} \frac{k^2 + z^2}{k^2 - z^2}\right]. \tag{22}$$

This formula is known as *the decomposition of* $\cot \pi z$ *into partial fractions*. On differentiating with respect to z, dividing by π and reversing the sign, we get *the decomposition of* $1/\sin^2 \pi z$ *into partial fractions*:

$$\frac{1}{\sin^2 \pi z} = \frac{1}{\pi^2}\left[\frac{1}{z^2} + 2\sum_{k=1}^{\infty} \frac{k^2+z^2}{(k^2-z^2)^2}\right],$$

or, noting that

$$2\frac{k^2+z^2}{(k^2-z^2)^2} = \frac{1}{(z+k)^2} + \frac{1}{(z-k)^2},$$

we can write the above in the more symmetrical form:

$$\frac{1}{\sin^2 \pi z} = \frac{1}{\pi^2}\sum_{k=-\infty}^{\infty} \frac{1}{(z-k)^2}. \tag{23}$$

Expression (22) leads to a striking *expansion of* $\cot z$ *into a power series*. On multiplying both sides by πz and replacing πz by z, i.e. z by z/π, we get:

$$z \cot z = 1 - \sum_{k=1}^{\infty} \frac{2z^2}{k^2\pi^2 - z^2}.$$

But

$$\frac{2z^2}{k^2\pi^2 - z^2} = \frac{2z^2}{k^2\pi^2\left(1 - \dfrac{z^2}{k^2\pi^2}\right)} =$$

$$= 2\frac{z^2}{k^2\pi^2}\left(1 + \frac{z^2}{k^2\pi^2} + \frac{z^4}{k^4\pi^4} + \ldots + \frac{z^{2n}}{k^{2n}\pi^{2n}} + \ldots\right) \quad (|z| < \pi).$$

On substituting this in the above formula and expanding in powers of z^2 we have

$$z \cot z = 1 - 2\frac{z^2}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2} - 2\frac{z^4}{\pi^4}\sum_{k=1}^{\infty}\frac{1}{k^4} - \ldots - 2\frac{z^{2n}}{\pi^{2n}}\sum_{k=1}^{\infty}\frac{1}{k^{2n}} - \ldots$$

Substitution of $z/2$ for z gives us:

$$\frac{1}{2}z \cot \frac{1}{2}z = 1 - \sum_{n=1}^{\infty}\left[\frac{2}{(2\pi)^{2n}}\sum_{k=1}^{\infty}\frac{1}{k^{2n}}\right]z^{2n}.$$

We shall denote the coefficient of z^{2n} by $B_n/(2n)!$:

$$\frac{1}{2}z \cot \frac{1}{2}z = 1 - \frac{B_1}{2!}z^2 - \frac{B_2}{4!}z^4 - \ldots - \frac{B_n}{(2n)!}z^{2n} - \ldots$$

$$B_n = \frac{2\cdot(2n)!}{(2\pi)^{2n}}\sum_{k=1}^{\infty}\frac{1}{k^{2n}}. \tag{24}$$

The first of the B_n can be found without difficulty by direct expansion of $(z/2)\cot z/2$, on writing this as the quotient of the series for $\cos z/2$ divided by the series for $\frac{\sin z/2}{z/2}$ [I, 130]:

$$B_1 = \frac{1}{6}, \quad B_2 = \frac{1}{30}, \quad B_3 = \frac{1}{42}, \quad B_4 = \frac{1}{30}, \quad B_5 = \frac{5}{66},$$

and it is clear at once that the B_n are rational numbers. They are known as Bernoulli numbers. On the other hand, knowing their values, we can find the sums of the series:

$$\sum_{k=1}^{\infty} \frac{1}{k^{2n}} = \frac{(2\pi)^{2n} B_n}{2\cdot(2n)!} \qquad (n = 1, 2, \ldots).$$

Occasionally, instead of Bernoulli numbers, we take the Eulerian numbers, defined by the expressions

$$A_0 = 1;\quad A_1 = -\frac{1}{2};\quad A_{2k} = \frac{(-1)^{k-1} B_k}{(2k)!};$$

$$A_{2k+1} = 0 \quad (k = 1, 2, 3, \ldots). \tag{25}$$

If we replace z by t/i in equation (24), the fact that

$$\frac{t}{2i}\cot\frac{t}{2i} = \frac{t}{2i}\cdot\frac{\cos\frac{t}{2i}}{\sin\frac{t}{2i}} = \frac{t}{2}\,\frac{e^{\frac{1}{2}t} + e^{-\frac{1}{2}t}}{e^{\frac{1}{2}t} - e^{-\frac{1}{2}t}} = \frac{t}{e^t - 1} + \frac{1}{2}t,$$

leads us to

$$\frac{t}{e^t - 1} = 1 - \frac{t}{2} + \frac{B_1 t^2}{2!} - \frac{B_2 t^4}{4!} + \ldots + (-1)^{n-1}\frac{B_n t^{2n}}{(2n)!} + \ldots =$$

$$= A_0 + A_1 t + A_2 t^2 + A_3 t^3 + \ldots$$

Bernoulli and Eulerian numbers are often encountered in various branches of analysis.

146. Periodic functions of period $2l$. It often becomes necessary to expand into a trigonometric series of cosines and sines a function defined in an interval $(-l, l)$ instead of $(-\pi, \pi)$; or alternatively, to expand in cosines or sines only a function defined in $(0, l)$.

This problem reduces to the above with the aid of a change of scale, i.e. by introducing in place of x an auxiliary variable ξ in accordance with

$$x = \frac{l\xi}{\pi};\ \xi = \frac{\pi x}{l}. \tag{26}$$

We put

$$f(x) = f\left(\frac{l\xi}{\pi}\right) = \varphi(\xi).$$

If $f(x)$ is defined in $(-l, l)$, $\varphi(\xi)$ is defined in the interval $(-\pi, \pi)$ of variation of ξ.

We obtain on expanding $\varphi(\xi)$ into a Fourier series:

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\xi + b_k \sin k\xi),$$

where, by (26):

$$\left.\begin{aligned} a_k &= \frac{1}{\pi}\int_{-\pi}^{+\pi} \varphi(\xi)\cos k\xi \,\mathrm{d}\xi = \frac{1}{\pi}\int_{-\pi}^{+\pi} f\left(\frac{l\xi}{\pi}\right)\cos k\xi\,\mathrm{d}\xi = \\ &= \frac{1}{l}\int_{-l}^{+l} f(x)\cos\frac{k\pi x}{l}\,\mathrm{d}x; \\ b_k &= \frac{1}{l}\int_{-l}^{+l} f(x)\sin\frac{k\pi x}{l}\,\mathrm{d}x. \end{aligned}\right\} \tag{27}$$

It follows that *Dirichlet's theorem remains valid for an interval* $(-l, l)$, *except that expansion* (6) *is replaced by the expansion*

$$\frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{k\pi x}{l} + b_k \sin\frac{k\pi x}{l}\right), \tag{28}$$

the coefficients a_k and b_k being defined in accordance with expressions (27).

The same applies to the cosine or sine expansions of a function $f(x)$ given in the interval $(0, l)$, the series obtained being

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\frac{k\pi x}{l}\,; \quad a_k = \frac{2}{l}\int_0^l f(x)\cos\frac{k\pi x}{l}\,\mathrm{d}x \tag{29}$$

and

$$\sum_{k=1}^{\infty} b_k \sin\frac{k\pi x}{l}\,; \quad b_k = \frac{2}{l}\int_0^l f(x)\sin\frac{k\pi x}{l}\,\mathrm{d}x. \tag{30}$$

Example. We find the sine expansion of the $f(x)$ defined by:

$$f(x) = \begin{cases} \sin\dfrac{\pi x}{l} \text{ for } 0 < x < \dfrac{1}{2}l \\ 0 \qquad \text{for } \dfrac{1}{2}l < x < l. \end{cases}$$

We have in this case:

$$b_k = \frac{2}{l}\int_0^l f(x)\sin\frac{k\pi x}{l}\,\mathrm{d}x = \frac{2}{l}\int_0^{\frac{l}{2}} \sin\frac{\pi x}{l}\sin\frac{k\pi x}{l}\,\mathrm{d}x,$$

since the integrand vanishes in $(l/2, l)$. Simple working, which we leave to the reader, gives:

$$b_k = \begin{cases} 0 & \text{for odd } k>1 \\ -\dfrac{(-1)^{\frac{1}{2}k}\,2k}{\pi(k^2-1)} & \text{for even } k \end{cases}$$

$$b_1 = \frac{1}{2},$$

so that

$$\frac{1}{2}\sin\frac{\pi x}{l} - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n n}{4n^2-1}\sin\frac{2n\pi x}{l} = \begin{cases} \sin\dfrac{\pi x}{l} & \text{for } 0 < x < \dfrac{1}{2}l \\ 0 & \text{for } \dfrac{1}{2}l < x < l \\ \dfrac{1}{2} & \text{for } x = \dfrac{1}{2}l \\ 0 & \text{for } x = 0 \text{ or } l. \end{cases} \tag{31}$$

The interval $(-l, l)$ can be replaced by any interval $(c, c + 2l)$ of length $2l$, as already mentioned as regards intervals of length 2π. With this, the sum of series (28) gives $f(x)$ in the interval $(c, c + 2l)$, and in evaluating the coefficients from formulae (27) the interval of integration $(-l, l)$ has to be replaced by the interval $(c, c + 2l)$.

147. Average error. The theory of Fourier series may be approached in a different manner. As above, let $f(x)$ be given in the interval $(-\pi, \pi)$. We form a linear combination of the first $(2n + 1)$ functions of family (4):

$$\frac{a_0}{2} + \sum_{k=1}^{n}(a_k \cos kx + \beta_k \sin kx), \tag{32}$$

where $a_0, a_1, \beta_1, \ldots, a_n, \beta_n$ are numerical coefficients. The expression written is generally known as a *trigonometric polynomial of the n-th order*. We consider the error resulting from taking the sum (32) for $f(x)$, i.e. we consider the difference:

$$\Delta_n(x) = f(x) - \left\{\frac{a_0}{2} + \sum_{k=1}^{n}(a_k \cos kx + \beta_k \sin kx)\right\}.$$

The greatest deviation Δ_n of sum (32) from $f(x)$ in the interval $(-\pi, \pi)$ is defined as the greatest value of $|\Delta_n(x)|$ in the interval. The smaller Δ_n, the more accurately the nth order trigonometric polynomial (32) represents $f(x)$. The quantity Δ_n is not suitable as a measure of the approximation, however, not only because the investigation of its value is difficult, but also because it is often more important in problems concerning the approximate representa-

tion of functions to achieve a reduction in the "average" or "probable" error rather than a reduction of the "greatest deviation". Figure 122 illustrates two approximate curves (dotted) for a given $f(x)$ (full curve). The greatest deviation of curve (1) is less than that of curve (2), yet in general (1) differs far more from the true curve than (2) does; the deviations of (2), though considerable in the interval $(-\pi, \pi)$, are of much smaller duration than those of (1).

When applying the method of least squares to checking the accuracy of a series of observations, we make use of the average or "root mean square" error, this being defined as follows: let the values obtained in measuring a quantity z be

$$z_1, z_2, \ldots, z_N;$$

the error of each measurement is

$$z - z_k \qquad (k = 1, 2, \ldots, N);$$

the average error δ_n is, by definition, given by

$$\delta_n^2 = \frac{1}{N} \sum_{k=1}^{N} (z - z_k)^2,$$

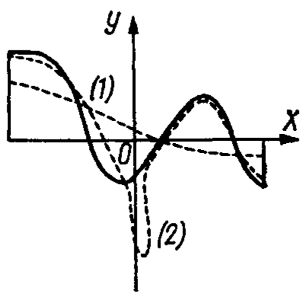

FIG. 122

i.e. δ_n *is the square root of the arithmetic mean of the squares of the errors.*

It is this average error that we take as a measure of the degree of approximation of sum (32) to our function $f(x)$. We only need to bear in mind that it is a matter of an infinite set, and not a finite number, of values, continuously distributed throughout the interval $(-\pi, \pi)$. Each separate error is thus in fact $\Delta(x)$, and the arithmetic mean of their squares will be

$$\frac{1}{2\pi} \int_{-\pi}^{+\pi} \Delta_n^2 (x)\, \mathrm{d}x,$$

whilst the average error δ_n of expression (32) is obtainable from the equation:

$$\delta_n^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \Delta_n^2 (x)\, \mathrm{d}x =$$

$$= \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left\{ f(x) - \frac{a_0}{2} - \sum_{k=1}^{n} (a_k \cos kx + \beta_k \sin kx) \right\}^2 \mathrm{d}x. \qquad (33)$$

It now only remains for us to select the constants $a_0, a_1, \beta_1, \ldots, a_n, \beta_n$ in such a way as to get the minimum δ_n^2, i.e. we have the ordinary problem of finding the minimum of the function δ_n^2 of $(2n+1)$ variables.

We first of all simplify expression (33) for δ_n^2. We find on squaring:

$$\left\{f(x) - \frac{a_0}{2} - \sum_{k=1}^{n}(a_k \cos kx + \beta_k \sin kx)\right\}^2 =$$

$$= [f(x)]^2 - a_0 f(x) - 2\sum_{k=1}^{n}(a_k \cos kx + \beta_k \sin kx) f(x) + \frac{a_0^2}{4} +$$

$$+ \sum_{k=1}^{n}(a_k^2 \cos^2 kx + \beta_k^2 \sin^2 kx) + \sigma_n, \tag{34}$$

where σ_n denotes a linear combination of expressions of the form:

$$\cos lx \cos mx, \quad \sin lx \sin mx \qquad (l \neq m),$$

$$\cos x \sin mx, \quad \cos lx, \quad \sin mx.$$

By the orthogonality of the trigonometric functions [143], the integrals of all these expressions vanish over the interval $(-\pi, \pi)$, whence it follows that the integral of σ_n over the interval vanishes. The integrals of $\cos^2 kx$ and $\sin^2 kx$ are equal to π, as we know, and we get on substituting (34) in (33):

$$\delta_n^2 = \frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)]^2\,dx - \frac{a_0}{2\pi}\int_{-\pi}^{+\pi} f(x)\,dx -$$

$$- \frac{1}{\pi}\sum_{k=1}^{n}\left[a_k \int_{-\pi}^{+\pi} f(x)\cos kx\,dx + \beta_k \int_{-\pi}^{+\pi} f(x)\sin kx\,dx\right] +$$

$$+ \frac{a_0^2}{4} + \frac{1}{2}\sum_{k=1}^{n}(a_k^2 + \beta_k^2).$$

We use expressions (9) for the coefficients of the Fourier series for $f(x)$ to rewrite the expression for δ_n^2 as follows:

$$\delta_n^2 = \frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)]^2\,dx - \frac{a_0 a_0}{2} - \sum_{k=1}^{n}(a_k a_k + \beta_k b_k) +$$

$$+ \frac{a_0^2}{4} + \frac{1}{2}\sum_{k=1}^{n}(a_k^2 + \beta_k^2),$$

or, on subtracting and adding the sum

$$\frac{a_0^2}{4} + \frac{1}{2} \sum_{k=1}^{n} (a_k^2 + b_k^2),$$

we can write our expression as:

$$\delta_n^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} [f(x)]^2 \, dx - \frac{a_0^2}{4} - \frac{1}{2} \sum_{k=1}^{n} (a_k^2 + b_k^2) + \frac{1}{4} (\alpha_0 - a_0)^2 +$$

$$+ \frac{1}{2} \sum_{k=1}^{n} \left[(\alpha_k - a_k)^2 + (\beta_k - b_k)^2 \right]. \tag{35}$$

The value of δ_n^2 is clearly a minimum when the last non-negative terms on the right-hand side vanish, i.e. when $\alpha_0 = a_0$ and generally, $\alpha_k = a_k$ and $\beta_k = b_k$ $(k = 1, 2, \ldots)$. It follows that *the average error of the approximation to a function $f(x)$ by means of an n-th order trigonometric polynomial is a minimum when the coefficients of the polynomial are the Fourier coefficients of $f(x)$.*

An important point must be noted. It follows from the result obtained that the values of α_k and β_k leading to minimum δ_n^2 do not depend on the subscript n. If we increase n, we must add new coefficients α_k and β_k; but those already calculated remain the same.

The least error ε_n is found from (35) by replacing α_k and β_k by a_k and b_k respectively:

$$\varepsilon_n^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} [f(x)]^2 \, dx - \frac{a_0^2}{4} - \frac{1}{2} \sum_{k=1}^{n} (a_k^2 + b_k^2), \tag{36}$$

or

$$2\varepsilon_n^2 = \frac{1}{\pi} \int_{-\pi}^{+\pi} [f(x)]^2 \, dx - \frac{a_0^2}{2} - \sum_{k=1}^{n} (a_k^2 + b_k^2). \tag{37}$$

On increasing the order n of the trigonometric polynomial, new negative (or at any rate, not positive) terms, $-a_{n+1}^2$, $-b_{n+1}^2$, $\ldots$, are added to the right-hand side of (37), and therefore *the error ε_n can only diminish on increasing n*, i.e. *the accuracy of the approximation increases (does not decrease) on increasing n.*

The quantity ε_n^2 is given by (33) if the α_k, β_k in it are replaced by a_k, b_k respectively, i.e. is given by the integral of the square of

a certain function, so that ε_n^2 is indeed positive or, more precisely not negative. On taking this into account, we get by (37):

$$\frac{a_0^2}{2} + \sum_{k=1}^{n} (a_k^2 + b_k^2) \leqslant \frac{1}{\pi} \int_{-\pi}^{+\pi} [f(x)]^2 \, dx. \tag{38}$$

We have so far made no explicit assumptions regarding $f(x)$. It is necessary for the above arguments that all the integrals employed should exist, i.e. that the Fourier coefficients should be calculable in accordance with formulae (9) and that the integral of the square of the function exists. We shall assume for definiteness that $f(x)$ is continuous or has a finite number of discontinuities of the first kind. All the integrals concerned certainly have a meaning with this assumption [I, 116]. We could in fact make far more general assumptions regarding $f(x)$ and in any case, those that figured in the Dirichlet conditions have no actual part in the above and future arguments. To return to inequality (38): as n increases, the sum of the positive terms on the left increases (does not diminish), whilst remaining less than the definite positive number occurring on the right. It follows immediately that the infinite series

$$\sum_{k=1}^{\infty} (a_k^2 + b_k^2)$$

is convergent [I, 120]. On letting n tend to infinity and passing to the limit in (38), we get:

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty} (a_k^2 + b_k^2) \leqslant \frac{1}{\pi} \int_{-\pi}^{+\pi} [f(x)]^2 \, dx. \tag{39}$$

On recalling that the general term of a convergent series must tend to zero as we move away from the initial term, we can state the following theorem:

THEOREM. *With the assumptions made regarding $f(x)$, its Fourier coefficients a_k and b_k tend to zero as $k \to +\infty$.*

The following is a fundamental problem from our new point of view: will the error ε_n tend to zero with indefinite increase of n? If we pass to the limit with n increasing indefinitely on the right-hand side of (37), we get instead of the finite sum $\sum\limits_{k=l}^{n}$ the infinite sum $\sum\limits_{k=l}^{\infty}$ i.e.

$$\lim_{n\to\infty} 2\varepsilon_n^2 = \frac{1}{\pi} \int_{-\pi}^{\pi} [f(x)]^2 \, dx - \frac{a_0^2}{2} - \sum_{k=1}^{\infty} (a_k^2 + b_k^2),$$

whence it follows that the fact that ε_n tends to zero is equivalent to our taking the sign of equality in (39), i.e.

$$\frac{1}{\pi}\int_{-\pi}^{+\pi}[f(x)]^2\,\mathrm{d}x = \frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2). \tag{40}$$

This is generally known as the *closure equation.* We show in the next section that $\varepsilon_n \to 0$, i.e. equation (40) is in fact valid for all functions $f(x)$ with the above-mentioned properties.

148. General orthogonal systems of functions. Most of the discussion of this chapter is based on the orthogonality of the functions of system (4) and not on the properties *per se* of the trigonometric functions. The discussion is on this account applicable to any system of orthogonal functions. Such systems are of frequent occurrence in mathematical physics, as we shall see. Let a system of real functions be given in the interval $a \leqslant x \leqslant b$, the functions being assumed continuous for the sake of clarity:

$$\varphi_1(x),\ \varphi_2(x),\ \ldots,\ \varphi_n(x),\ \ldots \tag{41}$$

We shall suppose that none of these functions is identically zero. The functions of system (41) are said to be orthogonal if

$$\int_a^b \varphi_m(x)\,\varphi_n(x)\,\mathrm{d}x = 0 \text{ with } m \neq n. \tag{42}$$

The integral of the square of each function of system (41) is equal to a certain positive constant. We introduce the following notation for these constants:

$$k_n = \int_a^b [\varphi_n(x)]^2\,\mathrm{d}x. \tag{43}$$

If we multiply each of the $\varphi_n(x)$ by the respective numerical factor $1/\sqrt{k_n}$, we obtain new functions

$$\psi_1(x) = \frac{1}{\sqrt{k_1}}\varphi_1(x);\quad \psi_2(x) = \frac{1}{\sqrt{k_2}}\varphi_2(x);\ \ldots,$$

$$\psi_n(x) = \frac{1}{\sqrt{k_n}}\varphi_n(x);\ \ldots,$$

which, by (42) and (43), satisfy not only the condition for orthogonality, but also the condition for the integral of the square of a function to be unity, i.e.

$$\int_a^b \psi_m(x)\,\psi_n(x)\,\mathrm{d}x = \begin{cases} 0 & \text{for } m \neq n \\ 1 & \text{for } m = n. \end{cases} \tag{44}$$

The functions

$$\psi_1(x),\ \psi_2(x),\ \ldots,\ \psi_n(x),\ \ldots \tag{45}$$

of a system are said to be *orthogonal and normalized* if they satisfy condition (44). Let $f(x)$ be a function defined in the interval (a, b) and let us suppose that it can be expanded in the interval in a series of functions (45):

$$f(x) = \sum_{k=1}^{\infty} c_k \psi_k(x), \tag{46}$$

where the c_k are numerical coefficients. We multiply both sides of (46) by $\psi_n(x)$ $(n = 1, 2, \ldots)$ and integrate over (a, b), making the assumption that the series on the right can be integrated term by term:

$$\int_a^b f(x)\,\psi_n(x)\,\mathrm{d}x = \sum_{k=1}^{\infty} c_k \int_a^b \psi_k(x)\,\psi_n(x)\,\mathrm{d}x.$$

On taking (44) into account, we get the following expressions for the coefficients c_n:

$$c_n = \int_a^b f(x)\,\psi_n(x)\,\mathrm{d}x \qquad (n = 1, 2, \ldots). \tag{47}$$

The c_n defined by these expressions are usually called the *generalized Fourier coefficients of the function $f(x)$ with respect to the system of functions* (45). The above remarks are only of an introductory nature, as in [142], and a rigorous treatment implies the following problem: if the coefficients c_n, calculated in accordance with formulae (47), are substituted in the series on the right-hand side of (46), will the series be convergent in the interval (a, b), and if so, will its sum be equal to $f(x)$? The solution of this problem naturally implies making certain assumptions regarding the properties of $f(x)$. The series obtained by substituting for the c_n their values from (47) is usually known at the *generalized Fourier series for $f(x)$*.

We can take an alternative approach by writing an expression for the average error in representing the given $f(x)$ as a finite sum of the form:

$$\sum_{k=1}^{n} \gamma_k \psi_k(x).$$

The square of the error will be given by:

$$\delta_n^2 = \frac{1}{b-a} \int_a^b \Big[f(x) - \sum_{k=1}^{n} \gamma_k \psi_k(x)\Big]^2 \mathrm{d}x.$$

If we take (44) and (47) into account and carry out working similar to that of [147], we get:

$$(b-a)\,\delta_n^2 = \int_a^b [f(x)]^2 \mathrm{d}x - \sum_{k=1}^{n} c_k^2 + \sum_{k=1}^{n} (\gamma_k - c_k)^2.$$

Hence it follows at once that δ_n^2 is a minimum when the γ_k are equal to the Fourier coefficients of $f(x)$, and denoting the minimum by ε_n, we have

$$(b-a)\,\varepsilon_n^2 = \int_a^b [f(x)]^2 \mathrm{d}x - \sum_{k=1}^{n} c_k^2.$$

Hence the convergence follows, as above, of the series

$$\sum_{k=1}^{\infty} c_k^2$$

and we have the inequality

$$\sum_{k=1}^{\infty} c_k^2 \leqslant \int_a^b [f(x)]^2 \mathrm{d}x, \tag{48}$$

which is generally known as *Bessel's inequality*. A basic problem here is whether ε_n tends to zero on indefinite increase of n, this being equivalent to having the equals sign in (48), i.e.

$$\int_a^b [f(x)]^2 \mathrm{d}x = \sum_{k=1}^{\infty} c_k^2. \tag{49}$$

This is known as the *closure equation for* $f(x)$ *with respect to the system of functions* (45). The system is said to be *closed* if equation (49) is valid for any continuous function $f(x)$ and for any function with a finite number of discontinuities of the first kind. It may be mentioned that if this is the case, it can be shown that (49) is valid, in fact, for a far wider class of functions.

The proof of the closure equation was given for various systems of orthogonal functions by V. A. Steklov, who pointed out the importance of the equation in the theory of orthogonal functions. A. M. Lyapunov first proved the equation for the case of trigonometric series.

We return to the system:

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots,\ \cos nx,\ \sin nx,\ \ldots$$

These functions possess the property of orthogonality in the interval $(-\pi, \pi)$ but they are not normalized, i.e. the integrals of their squares are not equal to unity. It follows from the discussion above [142] that the normalized orthogonal system is here

$$\frac{1}{\sqrt{2\pi}},\ \frac{1}{\sqrt{\pi}}\cos x,\ \frac{1}{\sqrt{\pi}}\sin x,\ \ldots,$$
$$\frac{1}{\sqrt{\pi}}\cos nx,\ \frac{1}{\sqrt{\pi}}\sin nx,\ \ldots$$

There is a simple geometrical analogy to the above. We take ordinary three-dimensional space and let **A** be a vector with components A_x, A_y, A_z along Cartesian axes. The square of the length of the vector is given by [103]:

$$|\mathbf{A}|^2 = A_x^2 + A_y^2 + A_z^2. \tag{50}$$

If we take two vectors **A** and **B**, the condition for them to be perpendicular is [103]:

$$A_x B_x + A_y B_y + A_z B_z = 0. \tag{51}$$

We now consider a far more complicated vector space: we take as a vector every real function $f(x)$, given in the interval (a, b), and possessing certain general properties similar to those discussed in previous articles, which enable the necessary integrations to be carried out.

By analogy with (50), we take the magnitude of the integral

$$\int_a^b [f(x)]^2\, dx$$

as the square of the length of the vector $f(x)$ of our function space, and by analogy with (51), we say that two vectors $f_1(x)$ and $f_2(x)$ of the space are perpendicular or orthogonal if

$$\int_b^b f_1(x) f_2(x)\, dx = 0.$$

We have here replaced the finite sums of (50) and (51) by integrals over (a, b). Using this terminology, we can say that condition (44) is equivalent to the fact that the vectors $\psi_n(x)$ appearing in family (45) consist of parallel pairs and are of unit length, i.e. vectors $\psi_n(x)$ in our function space are analogous to a system of mutually orthogonal unit vectors of ordinary space [102]. Let $f(x)$ be any vector of the function space. We can say that the c_n evaluated from (47) are the components of the vector $f(x)$ along the fundamental set $\psi_n(x)$. Bessel's inequality (48) is equivalent to the fact that the sum of the squares of the components does not exceed the square of the length of the vector itself. If we take a fundamental set of three mutually orthogonal unit vectors in three-dimensional space, we always have the sign of equality, by (50). But if we forget about the third unit vector, for instance (directed along the z axis), we now have to write instead of the equals sign:

$$A_x^2 + A_y^2 \leqslant |\mathbf{A}|^2,$$

the equality being only applicable to vectors lying in the XY plane, whilst a strict inequality applies for the rest. An infinite set of mutually orthogonal fundamental vectors exists in functional space and no simple verification is possible of the fact that none have been passed over.

If condition (49) for closure applies for all $f(x)$, i.e. for all vectors of the functional space, the condition, which is analogous to (50), also acts as a test for the fact that no fundamental vector is left out, i.e. for the fact that no new fundamental vector $\psi_0(x)$ can be added to system (45), orthogonal to all those already present. Let such a $\psi_0(x)$ in fact exist, i.e.

$$\int_a^b \psi_0(x)\,\psi_n(x)\,\mathrm{d}x = 0 \qquad (n = 1, 2, \ldots).$$

It follows from this, by (47), that all the Fourier coefficients of $\psi_0(x)$ with respect to functions (45) are zero. By hypothesis, (49) must be valid for all $f(x)$ and, in particular, for $\psi_0(x)$. But all the c_n are zero for the latter, and (49) gives:

$$\int_a^b [\psi_0(x)]^2\,\mathrm{d}x = 0. \tag{52}$$

If it is assumed say that $\psi_0(x)$ is continuous, it follows from (52) that $\psi_0(x)$ is identically zero in the interval $a \leqslant x \leqslant b$. The system (45) of orthogonal functions is said to be complete with respect to continuous functions if no continuous function exists, apart from those identically zero, orthogonal to all functions (45). It follows from what has been said that closure implies completeness with respect to continuous functions. The ability to obtain closure from completeness is bound up with a more general definition of the latter (not only with respect to continuous functions).

149. Practical harmonic analysis. The operation of expanding a given function in a Fourier series is called harmonic analysis. If the function $f(x)$ is given analytically, the problem is solved by means of formulae (27), which define the Fourier coefficients. In most practical cases, however, the function is given empirically, and the task of harmonic analysis is then to work out the most suitable methods either for evaluating the Fourier coefficients or for obtaining directly the harmonics of various orders in the given function.

Computational methods of harmonic analysis are based on the application of approximation formulae for integrals to the integrals for a_k and b_k. The rectangle formula is the simplest of these [I, 108].

We shall take an interval of length 2π, as is always possible by a suitable choice of scale on the x axis. We take $x = 0$ and $x = 2\pi$ as the ends of the interval. We now divide $(0, 2\pi)$ into n equal parts and denote the abscissae of the points of subdivision by

$$x_0 = 0,\ x_1,\ x_2,\ \ldots,\ x_{n-1},\ x_n = 2\pi,$$

the values of $f(x)$ at these points being denoted by

$$Y_0,\ Y_1,\ Y_2,\ \ldots,\ Y_{n-1},\ Y_n.$$

We now have by the rectangle formula:

$$a_k \sim \frac{2}{n} \sum_{i=0}^{n-1} Y_i \cos kx_i; \qquad b_k \sim \frac{2}{n} \sum_{i=0}^{n-1} Y_i \sin kx_i, \tag{53}$$

and the various methods for evaluating the coefficients a_k and b_k have the aim of simplifying formulae (53) and reducing to the minimum the number of multiplications necessary.

The method below, which is borrowed from W. Lohmann,† is based on certain transformations of formulae (53) and is extremely convenient both on

† *Harmonische Ánalyse zum Selqstunterricht*, Berlin, 1921.

account of the simplicity of the manipulations it requires and due to the high degree of accuracy of the results.

1. We take the graph of the curve to be analysed and draw the abscissa as close underneath it as possible (Fig. 123) in order to avoid both negative and very large ordinates. We divide the period of the curve into twenty equal parts.

2. We make up a table on a sheet ruled into squares as shown in Fig. 124. The numbers 1, 2, . . ., 20 in the first column denote the abscissae, whilst those in the second column denote the corresponding ordinates of the curve, obtained directly from the graph. It is useful to take the scale of the graph small enough for the ordinates to be integers.

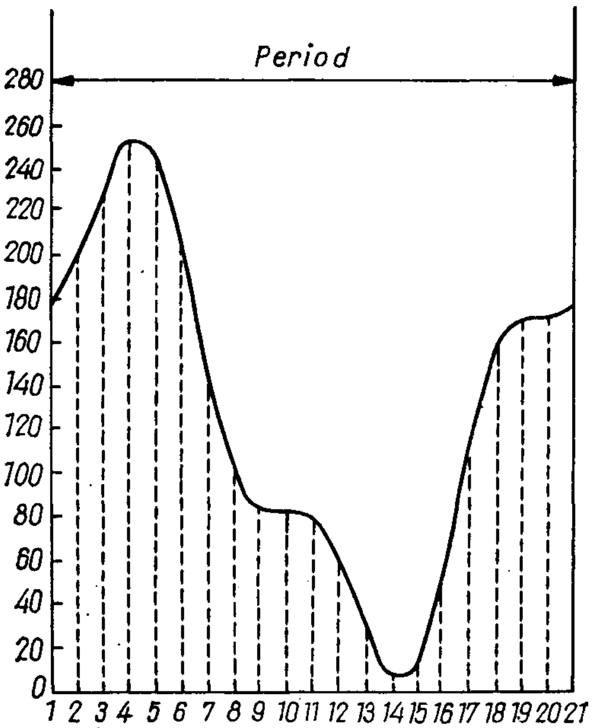

FIG. 123

1	175					
2	196	186	159	116	59	
3	230	218	186	136	69	
4	253	240	205	149	76	
5	245	233	198	145	74	
6	205					
7	135	128	109	80	41	
8	100	95	81	59	30	
9	82	78	66	48	25	
10	85	81	69	50	26	
11	82					
12	64	61	52	38	19	
13	29	28	23	17	9	
14	10	10	8	6	3	
15	15	14	12	9	5	
16	50					
17	110	105	89	65	33	
18	158	150	125	93	47	
19	174	165	141	103	52	
20	173	164	140	102	52	

FIG. 124

We write in the next column the products of the ordinates with $\cos 18° = 0.95$, in the next their products with $\cos 36° = 0.81$, in the next their products with $\cos 54° = 0.59$, and finally their products with $\cos 72° = 0.30$. We leave the last column blank, after shading its upper extremity in black (it is best to use arithmetic for the multiplications).

3. To find the constant term of the expansion $a_0/2 = r_0$, we add all the ordinates and divide the sum by twenty.

4. To determine the coefficients a_k, b_k $(k = 1, 2, \ldots, 10)$, a template of transparent waxed paper is prepared for each separate coefficient to patterns as shown in Figs. 125. The size of the squares on the template and of the template

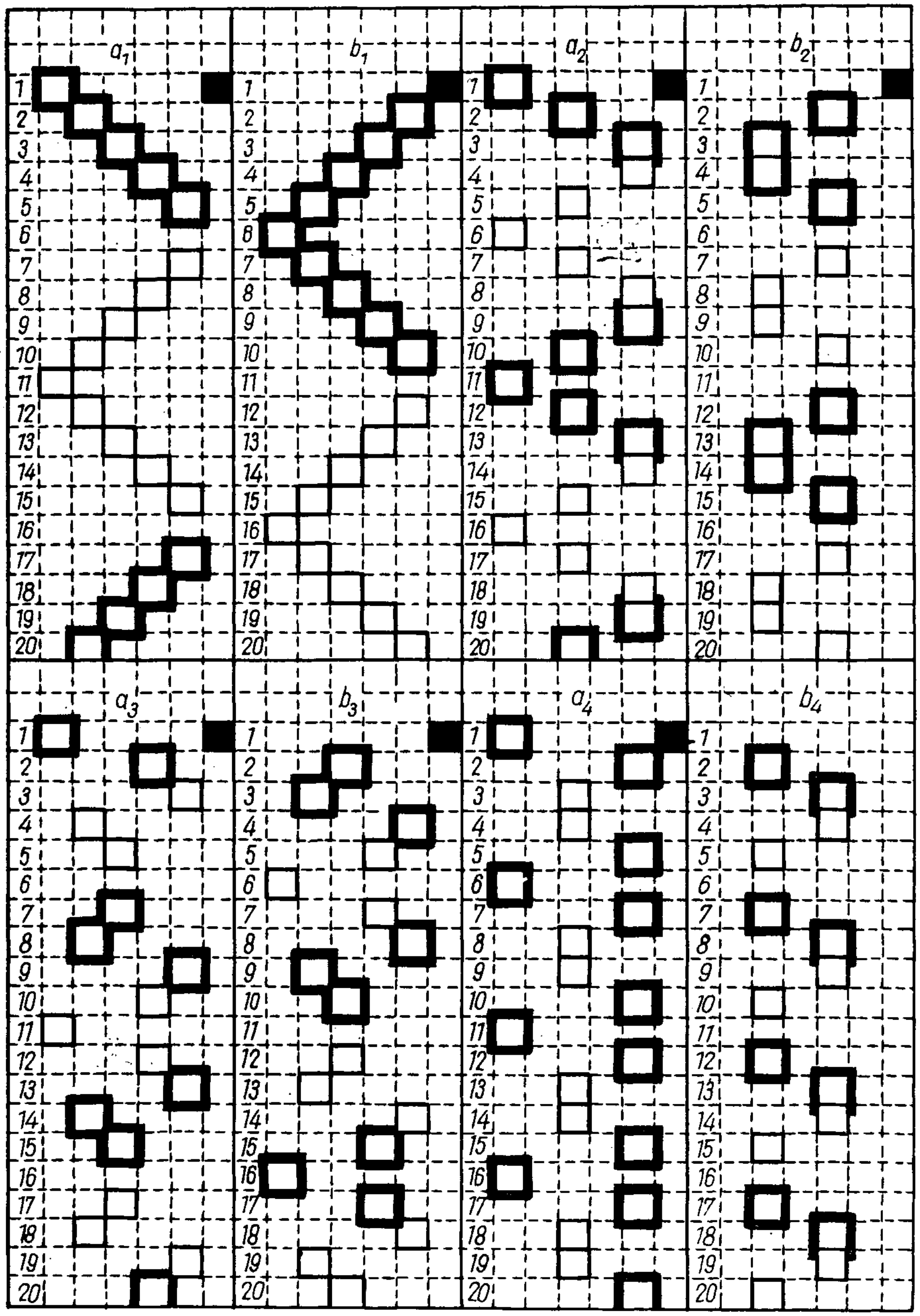

FIG. 125-a

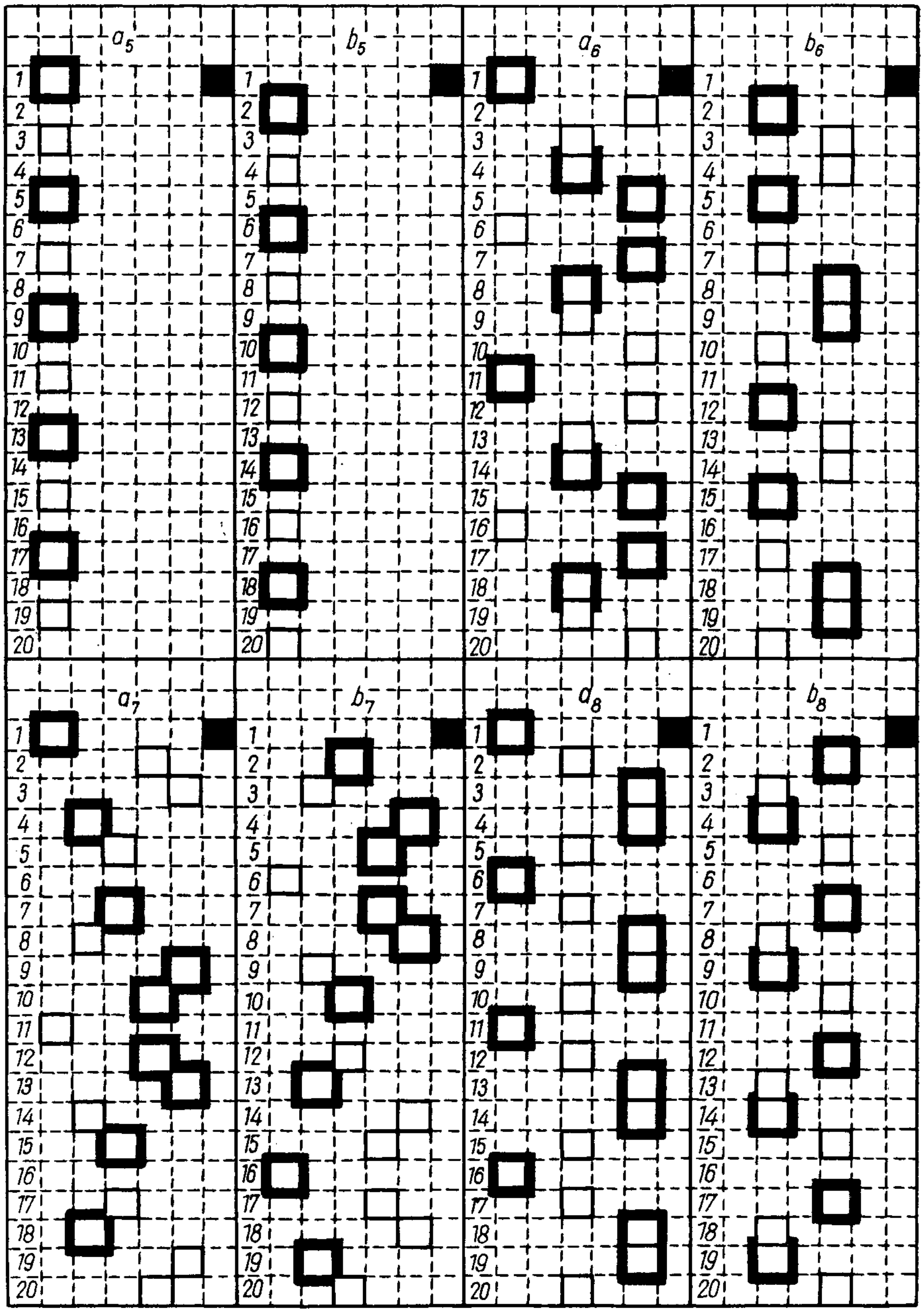

FIG. 125-b

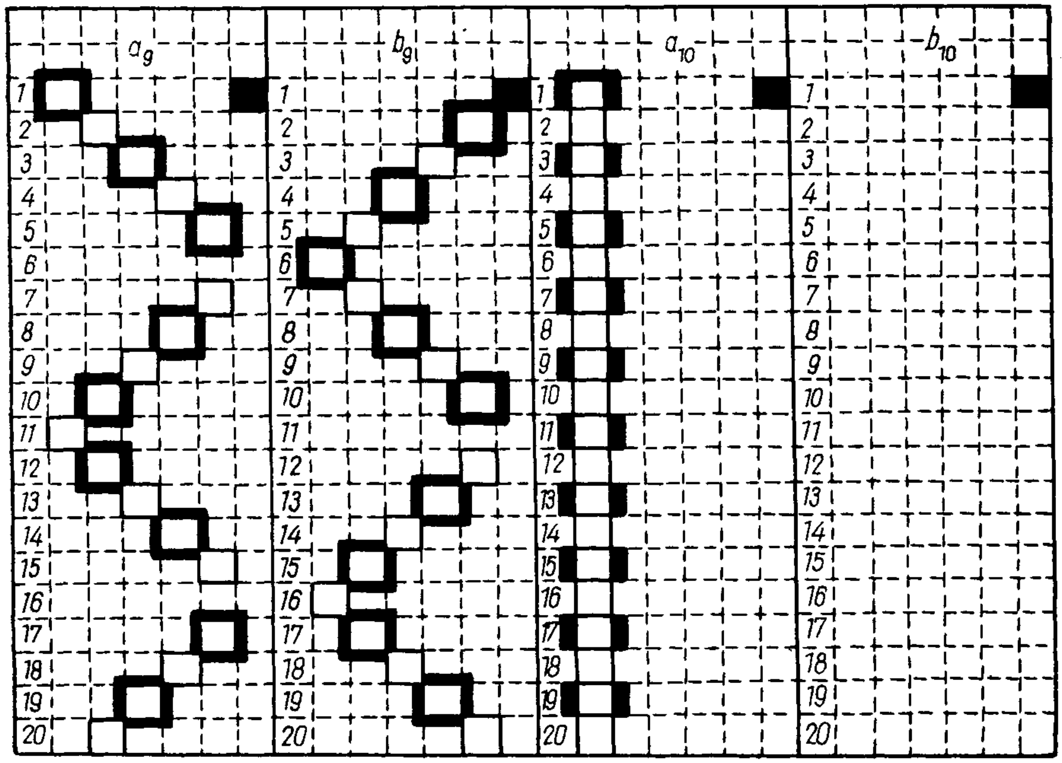

FIG. 125-c

itself must accurately correspond with those of the table of Fig. 124. The squares need to be bordered with thin and thick lines, or lines of different colours. Each template is laid on the table (Fig. 124) and the sum $\Sigma_{(+)}$ calculated of the numbers occupying squares with thick borders and similarly the sum $\Sigma_{(-)}$ of the numbers occupying squares with thin borders.

After carrying out this operation for all the templates, we find the differences between corresponding sums $(\Sigma_{(+)} - \Sigma_{(-)})$ and divide each difference by 10; the quotients obtained in fact give the coefficients $a_1, b_1, a_2, b_2, \ldots, a_{10}, b_{10}$.

5. We find the amplitudes $r_1, r_2, \ldots, r_{10}$ of the various harmonics of the required expansion from the formula:

$$r_k = \sqrt{a_k^2 + b_k^2}.$$

We determine the phases $\varphi_1, \varphi_2, \ldots, \varphi_9, \varphi_{10}$ of the harmonics from the formula:

$$\tan \varphi_k = \frac{a_k}{b_k}.$$

The angles φ_k can be found to an accuracy of $1°$ by making use of the table given below.

The two numbers are sought in the table between which the ratio $| a_k/b_k |$ lies, then the corresponding angle ψ_k is read off, in tens of degrees on the left of the relevant horizontal row and in degrees at the top of the relevant column.

		0° ↑	1° ↑	2° ↑	3° ↑	4° ↑	5° ↑	6° ↑	7° ↑	8° ↑	9° ↑
0°	0	0.01	0.03	0.04	0.06	0.08	0.10	0.11	0.13	0.15	0.17
10°	0.17	0·19	0.20	0.22	0.24	0.26	0.28	0.30	0.32	0.33	0.35
20°	0.35	0.37	0.39	0.41	0.43	0.46	0.48	0.50	0.52	0.54	0.57
30°	0.57	0.59	0.61	0.64	0.66	0.69	0.71	0.74	0.77	0.80	0.82
40°	0.82	0.85	0.88	0.92	0.95	0.98	1.02	1.05	1.09	1.13	1.17
50°	1.17	1.21	1.26	1.30	1.35	1.40	1.46	1.51	1.57	1.63	1.70
60°	1.70	1.75	1.84	1.90	2.06	2.10	2.20	2.30	2.40	2.5	2.7
70°	2.7	2.8	3.0	3.2	3.4	3.6	3.9	4.2	4.5	4.9	5.4
80°	5.4	6.0	6.7	7.6	9	10	13	16	23	38	115
90°	115	—	—	—	—	—	—	—	—	—	∞

Having found ψ_k, we get φ_k from the following table, according to the signs of a_k and b_k.

a_k	b_k	k
+	+	$\varphi_k = \psi_k$
+	—	$\varphi_k = 180° - \psi_k$
—	—	$\varphi_k = 180° + \psi_k$
—	+	$\varphi_k = 360° - \psi_k$

All these computations can usefully be set out in the form of a table, as shown below for the case of the curve of Fig. 123.

It may be noted in conclusion that the example worked out gives a reasonably accurate result only for the first harmonics.

$r_0 = 129$	$\Sigma_{(+)}$	$\Sigma_{(-)}$	$a = \frac{\Sigma_{(+)} - \Sigma_{(-)}}{10}$	$\Sigma_{(+)}$	$\Sigma_{(-)}$	$b = \frac{\Sigma_{(+)} - \Sigma_{(-)}}{10}$	a^2	b^2	$r = \sqrt{a^2 + b^2}$	$\tan\psi = \frac{b}{a}$	ψ°	φ°
1	1201	424	+77.7	1121	496	+62.5	6037.3	3906.3	100	1.24	51	51
2	832	819	+ 1.3	804	785	+ 1.9	1.7	3.6	2	0.68	34	34
3	653	968	−31.5	754	865	−11.1	992.3	123.2	33.4	2.84	71	251
4	821	838	− 1.7	785	798	− 1.3	2.9	1.7	2	1.38	54	234
5	641	634	+ 0.5	654	640	+ 1.4	0.3	2.0	2	0.36	20	20
6	832	827	+ 0.5	797	786	+ 1.1	0.3	1.2	1	0.45	24	24
7	808	813	− 0.5	802	817	− 1.5	0.3	2.3	2	0.33	18	198
8	823	828	− 0.5	792	797	− 0.5	0.3	0.3	1	1.00	45	225
9	816	809	+ 0.7	815	802	+ 1.3	0.5	1.7	2	0.54	29	29
10	1277	1294	− 1.7	—	—	—	—	—	2	∞	90	270

§ 15. Supplementary remarks on the theory of Fourier series

150. Expansion in Fourier series. The present section is concerned with a more penetrating and rigorous account of the theory of Fourier series and starts with a proof of the theorem regarding the Fourier expansion of $f(x)$. We shall impose conditions on $f(x)$ different to the Dirichlet conditions [143] so as to simplify the proof. The proof of the Dirichlet theorem is given later.

We take the Fourier series for the function $f(x)$:

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \tag{1}$$

where

$$a_k = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \cos kt \, dt; \qquad b_k = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \sin kt \, dt,$$

the variable of integration being denoted by t so as to avoid confusion in later working with the variable x of expression (1). We substitute the expressions for a_k and b_k in (1) and find the sum of the first $(2n+1)$ terms of the Fourier series for $f(x)$, the sum being written $S_n(f)$:

$$S_n(f) = \frac{a_0}{2} + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx) =$$

$$= \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\Big[\frac{1}{2} \sum_{k=1}^{n}(\cos kt \cos kx + \sin kt \sin kx)\Big]\,dt =$$

$$= \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\Big[\frac{1}{2} + \sum_{k=1}^{n}\cos k(t-x)\Big]\,dt.$$

But we can write [I, 174]:

$$1 + \cos\varphi + \cos 2\varphi + \ldots + \cos(n-1)\varphi = \frac{\sin\left(n-\frac{1}{2}\right)\varphi + \sin\frac{\varphi}{2}}{2\sin\frac{\varphi}{2}}.$$

On replacing n by $(n+1)$ in this equation and subtracting a half from both sides, we get:

$$\frac{1}{2} + \cos\varphi + \cos 2\varphi + \ldots + \cos n\varphi = \frac{\sin\frac{(2n+1)\varphi}{2}}{2\sin\frac{\varphi}{2}},$$

whence

$$\frac{1}{2} + \sum_{k=1}^{n}\cos k(t-x) = \frac{\sin\frac{(2n+1)(t-x)}{2}}{2\sin\frac{t-x}{2}}, \tag{2}$$

so that we can write the previous expression for $S_n(f)$ in the form:

$$S_n(f) = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\,\frac{\sin\frac{(2n+1)(t-x)}{2}}{2\sin\frac{t-x}{2}}\,dt.$$

We make a periodic continuation of $f(x)$, assigned in the interval $(-\pi, \pi)$, so that it can be assumed to have period 2π and to be defined for all real x. The fraction appearing under the integral sign also has period 2π with respect to t, by virtue of (2). If we take into consideration the "remark" in [142], we can replace the interval

of integration $(-\pi, \pi)$ in the above integral by any interval of length 2π. We take any value x of the independent variable and use the interval of integration $(x - \pi, x + \pi)$:

$$S_n(f) = \frac{1}{\pi} \int\limits_{x-\pi}^{x+\pi} f(t) \frac{\sin \frac{(2n+1)(t-x)}{2}}{2 \sin \frac{t-x}{2}} \, dt .$$

It must be repeated that throughout the future working $f(x)$ is understood to be continued outside $(-\pi, \pi)$ for all real x in the manner described above.

We split the integral into two: firstly $\int\limits_{x-\pi}^{x}$ and secondly $\int\limits_{x}^{x+\pi}$. We introduce new variables of integration z instead of t, given by $t = x - 2z$ as regards the first integral, and by $t = x + 2z$ as regards the second. On changing the variables in the integrands and putting in the new limits of integration, we get:

$$S_n(f) = \frac{1}{\pi} \int\limits_0^{\frac{\pi}{2}} f(x - 2z) \frac{\sin (2n+1) z}{\sin z} \, dz + $$

$$+ \frac{1}{\pi} \int\limits_0^{\frac{\pi}{2}} f(x + 2z) \frac{\sin (2n+1) z}{\sin z} \, dz. \tag{3}$$

If we suppose that $f(x)$ is equal to unity throughout $(-\pi, \pi)$, the free term $a_0/2$ of its Fourier expansion is evidently unity whilst the remaining terms are zero, i.e. $S_n(f)$ is unity for any n, and we have the following equation:

$$1 = \frac{2}{\pi} \int\limits_0^{\frac{\pi}{2}} \frac{\sin (2n+1)z}{\sin z} \, dz \quad (n = 1, 2, 3 \ldots). \tag{4}$$

We prove a lemma before turning to the fundamental theorem:

Lemma. *If the interval (a, b) is all or part of $(-\pi, \pi)$ and $\psi(z)$ is continuous, or has a finite number of discontinuities of the first kind, in (a, b), the integrals*

$$\frac{1}{\pi} \int\limits_a^b \psi(z) \cos nz \, dz \quad \text{and} \quad \frac{1}{\pi} \int\limits_a^b \psi(z) \sin nz \, dz$$

tend to zero on indefinite increase of the integer n. If (a, b) is $(-\pi, \pi)$, this lemma is precisely the same as the theorem of [147]. We now let (a, b) be part of $(-\pi, \pi)$. We continue $\psi(z)$ throughout the longer interval $(-\pi, \pi)$ by making it zero in the parts of this interval outside (a, b), i.e. we define a new function $\psi_1(z)$ such that $\psi_1(z) = \psi(z)$ for $a \leqslant z \leqslant b$ and $\psi_1(z) = 0$ for z in $(-\pi, \pi)$ but outside (a, b). We can now write, for instance,

$$\frac{1}{\pi}\int\limits_a^b \psi(z)\cos nz\,dz = \frac{1}{\pi}\int\limits_{-\pi}^{+\pi}\psi_1(z)\cos nz\,dz,$$

and this integral tends to zero by the theorem of [147] referred to above. It may be pointed out that $\psi_1(z)$ is also continuous, or else has a finite number of discontinuities of the first kind in $(-\pi, \pi)$. The lemma is easily shown to remain valid if (a, b) is any finite interval.

We now turn to the basic theorem for the expansion of $f(x)$ in a Fourier series. As usual, we take $f(x)$ as either continuous or possessing a finite number of discontinuities of the first kind in the interval $-\pi \leqslant x \leqslant \pi$.

If we multiply both sides of equation (4) by the factor $f(x)$, take the factor under the integral sign and subtract the resulting equation from (3), we get:

$$S_n(f) - f(x) = \frac{1}{\pi}\int\limits_0^{\frac{\pi}{2}}[f(x-2z) - f(x)]\frac{\sin(2n+1)z}{\sin z}\,dz + \\ + \frac{1}{\pi}\int\limits_0^{\frac{\pi}{2}}[f(x+2z) - f(x)]\frac{\sin(2n+1)z}{\sin z}\,dz,$$

which we can also write in the form:

$$S_n(f) - f(x) = \frac{1}{\pi}\int\limits_0^{\frac{\pi}{2}}\frac{f(x-2z) - f(x)}{-2z}\cdot\frac{-2z}{\sin z}\sin(2n+1)z\,dz + \\ + \frac{1}{\pi}\int\limits_0^{\frac{\pi}{2}}\frac{f(x+2z) - f(x)}{2z}\cdot\frac{2z}{\sin z}\sin(2n+1)z\,dz. \qquad (5)$$

To prove that Fourier series (1) for $f(x)$ is convergent to the sum $f(x)$, we have to show that the difference $[S_n(f) - f(x)]$ tends to zero on indefinite increase of n.

We consider the function

$$\psi(z) = \frac{f(x - 2z) - f(x)}{-2z} \cdot \frac{-2z}{\sin z}$$

in the interval $(0, \pi/2)$. It can have discontinuities of the first kind originating from the discontinuities of $f(x - 2z)$ and furthermore, the value $z = 0$ must be specially investigated. We suppose that $f(z)$ is not only continuous at the point x that we have taken but also has a derivative. It follows from the definition of derivative and the obvious equation

$$\lim_{z \to 0} \frac{-2z}{\sin z} = -2,$$

that $\psi(z)$ tends to a definite limit, equal to $-2f'(x)$, as $z \to 0$. Hence the above lemma is applicable to $\psi(z)$, and the first term on the right-hand side of (5) tends to zero on indefinite increase of n. It can be shown in a similar manner that the second term tends to zero, whence it follows that the difference $[S_n(f) - f(x)]$ tends to zero at the point x. We thus obtain the following theorem:

THEOREM. *If $f(x)$ is continuous or has a finite number of discontinuities of the first kind in the interval $(-\pi, \pi)$, its Fourier series is convergent to the sum $f(x)$ at every point x at which $f(x)$ has a derivative.*

It is easy to obtain a more general result. We suppose that at the point x the function is continuous or even has a discontinuity of the first kind, whilst the finite limits exist:

$$\lim_{h \to +0} \frac{f(x - h) - f(x - 0)}{-h} \quad \text{and} \quad \lim_{h \to +0} \frac{f(x + h) - f(x + 0)}{h}. \tag{6}$$

The existence of these limits, i.e. of the derivatives to the left and right, is equivalent geometrically to the existence of a definite tangent to the left and to the right. We have in this case the following supplement to the theorem: *if the finite limits* (6) *exist, the Fourier series for $f(x)$ is convergent at this point to the sum $[f(x - 0) + f(x + 0)]/2$ (or to $f(x)$ if $f(x)$ is continuous).*

We can write on multiplying (4) by $[f(x-0)+f(x+0)]/2$ and subtracting from (3):

$$S_n(f) - \frac{f(x-0)+f(x+0)}{2} =$$

$$= \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \frac{f(x-2z)-f(x-0)}{-2z} \cdot \frac{-2z}{\sin z}\sin(2n+1)z\,dz +$$

$$+ \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \frac{f(x+2z)-f(x+0)}{2z} \cdot \frac{2z}{\sin z}\sin(2n+1)z\,dz. \qquad (7)$$

We have to prove that the right-hand side tends to zero on indefinite increase of n.

We can say from the existence of limits (6) that both the fractions

$$\frac{f(x-2z)-f(x-0)}{-2z} \quad \text{and} \quad \frac{f(x+2z)-f(x+0)}{2z}$$

have finite limits as $z \to 0$, and we can use exactly the same arguments as above to see that both integrals on the right of (7) tend to zero on indefinite increase of n. This proves the supplement to the theorem.

For $x = \pi$ and $x = -\pi$, by the periodic continuation of $f(x)$, limits (6) become:

$$\lim_{h\to+0} \frac{f(-\pi+h)-f(-\pi+0)}{h} \quad \text{and} \quad \lim_{h\to+0} \frac{f(\pi-h)-f(\pi-0)}{-h},$$

and the sum of the series will be:

$$\frac{f(-\pi+0)+f(\pi-0)}{2}.$$

It should be noticed that in all the examples considered in the previous section, $f(x)$ satisfies at all points the conditions of the above theorem and its supplement.

151. Second mean value theorem. For the proof of Dirichlet's theorem and a more detailed account of Fourier series we need a proposition of the integral calculus that has some similarity to the mean value theorem given in Vol. I [I, 95]. The new proposition is usually referred to as *the second mean value theorem*, and may be stated as follows: *if $\varphi(x)$ is monotonic and bounded and has a finite number of discontinuities in the finite interval $a \leqslant x \leqslant b$, whilst $f(x)$ is continuous, we have:*

$$\int_a^b \varphi(x) f(x)\,dx = \varphi(a+0)\int_a^{\xi} f(x)\,dx + \varphi(b-0)\int_{\xi}^b f(x)\,dx, \qquad (8)$$

where *ξ is a number belonging to the interval (a, b).*

It is sufficient to prove (8) for an increasing (non-decreasing) $\varphi(x)$, since evidently if $\varphi(x)$ is decreasing, $[-\varphi(x)]$ is increasing, and on applying (8) for $[-\varphi(x)]$ and changing the sign on both sides, we get (8) for $\varphi(x)$ itself. It can be shown moreover that it is sufficient to prove (8) for the case of $\varphi(a + 0) = 0$. We assume that (8) is proved for this case and that $\varphi(x)$ does not satisfy the condition stated. We introduce a new monotonic function $\psi(x) = \varphi(x) - \varphi(a + 0)$. The end values of this function will be $\psi(a + 0) = 0$ and $\psi(b - 0) = \varphi(b - 0) - \varphi(a + 0)$. Equation (8) is applicable to $\psi(x)$ by hypothesis, and gives, since $\psi(a + 0) = 0$:

$$\int_a^b \psi(x)\, f(x)\, \mathrm{d}x = \psi(b - 0) \int_\xi^b f(x)\, \mathrm{d}x,$$

or

$$\int_a^b [\varphi(x) - \varphi(a + 0)]\, f(x)\, \mathrm{d}x = [\varphi(b - 0) - \varphi(a + 0)] \int_\xi^b f(x)\, \mathrm{d}x,$$

whence

$$\int_a^b \varphi(x)\, f(x)\, \mathrm{d}x = \varphi(a + 0) \left[\int_a^b f(x)\, \mathrm{d}x - \int_\xi^b f(x)\, \mathrm{d}x \right] + \varphi(b - 0) \int_\xi^b f(x)\, \mathrm{d}x,$$

which leads immediately to equation (8) for $\varphi(x)$. All in all, therefore, it is sufficient to prove (8) for increasing, or more precisely, non-decreasing $\varphi(x)$ for which $\varphi(a + 0) = 0$. This type of function clearly has non-negative values in (a, b).

We carry out the proof by subdividing (a, b) with the aid of the points:

$$x_0 = a, x_1, x_2, \ldots, x_{i-1}, x_i \ldots, x_{n-1}, x_n = b.$$

We know from [I, 95] that:

$$\int_{x_{i-1}}^{x_i} f(x)\, \mathrm{d}x = f(\xi_i)\,(x_i - x_{i-1}),$$

where ξ_i belongs inside (x_{i-1}, x_i).

We form the sum:

$$\sum_{i=1}^{n} \varphi(\xi_i)\, f(\xi_i)\,(x_i - x_{i-1}) = \sum_{i=1}^{n} \varphi(\xi_i) \int_{x_{i-1}}^{x_i} f(x)\, \mathrm{d}x.$$

On indefinite increase of n and indefinite decrease of the greatest of the lengths of sub-intervals (x_{i-1}, x_i), this sum tends to the definite integral (as we know from Vol. I), i.e. we have:

$$\int_a^b \varphi(x)\, f(x)\, \mathrm{d}x = \lim \sum_{i=1}^{n} \varphi(\xi_i) \int_{x_{i-1}}^{x_i} f(x)\, \mathrm{d}x.$$

We now investigate the sum

$$\sum_{i=1}^{n} \varphi(\xi_i) \int_{x_{i-1}}^{x_i} f(x)\,dx = \sum_{i=1}^{n} \varphi(\xi_i) [\int_{x_{i-1}}^{b} f(x)\,dx - \int_{x_i}^{b} f(x)\,dx] =$$

$$= \varphi(\xi_1) \int_{a}^{c} f(x)\,dx + \sum_{i=2}^{n} [\varphi(\xi_i) - \varphi(\xi_{i-1})] \int_{x_{i-1}}^{a} f(x)\,dx. \tag{9}$$

The integrals

$$\int_{a}^{b} f(x)\,dx, \int_{x_1}^{b} f(x)\,dx, \int_{x_2}^{b} f(x)\,dx, \ldots, \int_{x_{i-1}}^{b} f(x)\,dx, \ldots, \int_{x_{n-1}}^{b} f(x)\,dx \tag{10}$$

represent particular values of the function

$$\int_{x}^{b} f(x)\,dx = -\int_{b}^{x} f(x)\,dx, \tag{11}$$

which is a continuous function of the limit of integration x [I, 96], so that all the values (10) lie between the least and greatest values m and M of function (11).

If we take into account the fact that in (9) all the factors

$$\varphi(\xi_1) \text{ and } \varphi(\xi_i) - \varphi(\xi_{i-1})$$

are non-negative, and replace integrals (10) on the right of (9) firstly by m, then by M, we get:

$$\sum_{i=1}^{n} \varphi(\xi_i) \int_{x_{i-1}}^{x_i} f(x)\,dx \geqslant \left\{\varphi(\xi_1) + \sum_{i=2}^{n} [\varphi(\xi_i) - \varphi(\xi_{i-1})]\right\} m = \varphi(\xi_n)\,m,$$

$$\sum_{i=1}^{n} \varphi(x_i) \int_{x_{i-1}}^{x_i} f(x)\,dx \leqslant \left\{\varphi(\xi_1) + \sum_{i=2}^{n} [\varphi(\xi_i) - \varphi(\xi_{i-1})]\right\} M = \varphi(\xi_n)\,M,$$

i.e.

$$\varphi(\xi_n)\,m \leqslant \sum_{i=1}^{n} \varphi(\xi_i) \int_{x_{i-1}}^{x_i} f(x)\,dx \leqslant \varphi(\xi_n)\,M,$$

whilst in the limit, as $n \to \infty$ and the greatest of the lengths of the (x_{i-1}, x_i) diminishes indefinitely, since we have

$$\xi_n \to b - 0 \text{ and } \varphi(\xi_n) \to \varphi(b - 0),$$

the inequality becomes:

$$\varphi(b-0)\,m \leqslant \int_{a}^{b} \varphi(x) f(x)\,dx < \varphi(b-0)\,M,$$

i.e.

$$\int_{a}^{b} \varphi(x) f(x)\,dx = \varphi(b-0)\,P,$$

where P is a number lying in the interval (m, M). But the continuous function (11) takes all values lying between its least value m and greatest value M in the interval (a, b) [I, 43], including the number P, so that it must in fact be possible to find a ξ in (a, b) such that

$$\int_{\xi}^{b} f(x)\,\mathrm{d}x = P,$$

which gives us

$$\int_{a}^{b} \varphi(x) f(x)\,\mathrm{d}x = \varphi(b-0)\int_{\xi}^{b} f(x)\,\mathrm{d}x,$$

and this is the same as (8) since $\varphi(a + 0) = 0$ by hypothesis. It may be mentioned that (8) can be proved without assuming continuity of $f(x)$ and a finite number of discontinuities of $\varphi(x)$, though we shall not dwell on this. Finally, it must be pointed out that the more general formula than (8) may be proved:

$$\int_{a}^{q} \varphi(x) f(x)\,\mathrm{d}x = A\int_{a}^{\xi} f(x)\,\mathrm{d}x + B\int_{\xi}^{b} f(x)\,\mathrm{d}x.$$

where the numbers A and B must satisfy the conditions $A < \varphi(a + 0)$ and $B \geqslant \varphi(b - 0)$.

Corollary. We saw in [147] that, given certain conditions, the Fourier coefficients a_n and b_n of $f(x)$ tend to zero as $n \to \infty$. If $f(x)$ satisfies Dirichlet conditions, it can be shown more precisely that for large n the coefficients are infinitesimals of order not lower than $1/n$, i.e. we can write:

$$|a_n| < \frac{M}{n}; \quad |b_n| < \frac{M}{n},$$

where M is a definite positive number. By hypothesis, the interval $(-\pi, \pi)$ can be divided into a finite number of parts in each of which $f(x)$ is monotonic and bounded. Let (α, β) be such a sub-interval. The coefficient a_n is the sum of a finite number of terms of the form:

$$\frac{1}{\pi}\int_{\alpha}^{\beta} f(x)\cos nx\,\mathrm{d}x,$$

which can be transformed by the mean value theorem:

$$\frac{1}{\pi}\int_{\alpha}^{\beta} f(x)\cos nx\,\mathrm{d}x = \frac{1}{\pi} f(\alpha+0)\int_{\alpha}^{\xi}\cos nx\,\mathrm{d}x + \frac{1}{\pi} f(\beta-0)\int_{\xi}^{\beta}\cos nx\,\mathrm{d}x =$$

$$= \frac{f(\alpha+0)(\sin n\xi - \sin n\alpha) + f(\beta-0)(\sin n\beta - \sin n\xi)}{\pi n}.$$

This gives us for each separate term in the expression for a_n an upper limit of the form M/n, where $M = 2\,|f(\alpha + 0)|/\pi + 2\,|f(\beta - 0)|/\pi$. There will clearly be a similar upper limit for the sum of a finite number of terms of this form, i.e. for a_n. Analogous reasoning applies in the case of b_n.

If $f(x)$ is continuous and $f(-\pi) = f(\pi)$, whilst the derivative $f'(x)$ exists and satisfies Dirichlet conditions, we obtain, on integrating by parts and noticing that the term outside the integral vanishes since $f(-\pi) = f(\pi)$:

$$nb_n = \frac{n}{\pi} \int_{-\pi}^{+\pi} f(x) \sin nx \, dx = -\frac{1}{\pi} \int_{-\pi}^{+\pi} f(x) \, d \cos nx = \frac{1}{\pi} \int_{-\pi}^{+\pi} f'(x) \cos nx \, dx.$$

But the last integral, being the Fourier coefficient of $f'(x)$ satisfying Dirichlet conditions, must have the same upper limit as above, so that with the assumptions made we get for b_n:

$$|b_n| \leqslant \frac{M}{n^2}.$$

We can write an analogous expression for a_n. A more detailed discussion follows later of the values of the Fourier coefficients in relation to the properties of $f(x)$.

152. Dirichlet integrals. It is clear from expression (3) that the question of the convergence of a Fourier series, i.e. of the existence of a limit of the sum $S_n(f)$, amounts to an investigation of integrals of the form:

$$\int_a^b \varphi(z) \frac{\sin mz}{\sin z} \, dz.$$

We shall consider the simpler type of integral:

$$\frac{1}{\pi} \int_a^b \varphi(z) \frac{\sin mz}{z} \, dz, \tag{12}$$

which is known as a *Dirichlet integral*. We prove the following lemma in regard to this:

LEMMA. *If $\varphi(z)$ satisfies Dirichlet conditions in the interval (a, b), we can say: (1) if $a = 0$ and $b > 0$, integral (12) tends on indefinite increase of m to the limit $\varphi(+0)/2$; (2) if $a = 0$ and $b < 0$, the limit becomes $\varphi(-0)/2$; (3) if $a < 0$ and $b > 0$, the limit is $\{\varphi(-0) + \varphi(+0)\}/2$; (4) if a and $b > 0$ or a and $b < 0$, the limit is zero.* It can easily be seen that it is sufficient to prove only the first statement, from which the remainder follow without difficulty. We prove statements 3 and 4, for instance, on the assumption that the first has been proved:

$$\frac{1}{\pi} \int_b^a \varphi(z) \frac{\sin mz}{z} \, dz = \frac{1}{\pi} \int_0^b \varphi(z) \frac{\sin mz}{z} \, dz - \frac{1}{\pi} \int_0^a \varphi(z) \frac{\sin mz}{z} \, dz.$$

If a and $b > 0$, both terms on the right-hand side here have the limit $\varphi(+0)/2$ and their difference consequently tends to zero, which proves statement 4.

If $a < 0$ and $b > 0$, we replace the variable of integration z by $(-z)$ in the second term on the right, and obtain:

$$\frac{1}{\pi}\int_a^b \varphi(z)\,\frac{\sin mz}{z}\,\mathrm{d}z = \frac{1}{\pi}\int_0^b \varphi(z)\,\frac{\sin mz}{z}\,\mathrm{d}z + \frac{1}{\pi}\int_0^{-a} \varphi(-z)\,\frac{\sin mz}{z}\,\mathrm{d}z.$$

Since b and $(-a) > 0$, we can apply statement 1 to both integrals and obtain:

$$\frac{1}{\pi}\int_0^b \varphi(z)\,\frac{\sin mz}{z}\,\mathrm{d}z \to \frac{1}{2}\,\varphi(+0) + \frac{1}{2}\,\varphi(-0) = \frac{\varphi(-0)+\varphi(+0)}{2}\;.$$

We now turn to the proof of statement 1, i.e. we prove that with $b > 0$,

$$\frac{1}{\pi}\int_0^b \varphi(z)\,\frac{\sin mz}{z}\,\mathrm{d}z \to \frac{1}{2}\,\varphi(+0). \tag{13}$$

We shall assume for the present in our proof that $\varphi(z)$ not only satisfies Dirichlet conditions but is also monotonic in $(0, b)$.

We had earlier the result:

$$\int_0^\infty \frac{\sin x}{x}\,\mathrm{d}x = \frac{\pi}{2}\;. \tag{14}$$

We consider the integral:

$$\int_0^c \frac{\sin x}{x}\,\mathrm{d}x.$$

This is a continuous function of c, vanishing for $c = 0$ and tending to $\pi/2$ as $c \to +\infty$. We can infer from this that for all positive c the integral written remains less in absolute value than a definite positive number M. We now take the integral with two positive limits

$$\int_a^b \frac{\sin x}{x}\,\mathrm{d}x. \tag{15}$$

We clearly have

$$\int_a^b \frac{\sin x}{x}\,\mathrm{d}x = \int_0^b \frac{\sin x}{x}\,\mathrm{d}x - \int_0^a \frac{\sin x}{x}\,\mathrm{d}x$$

and

$$\left|\int_a^b \frac{\sin x}{x}\,\mathrm{d}x\right| \leqslant \left|\int_0^b \frac{\sin x}{x}\,\mathrm{d}x\right| + \left|\int_0^a \frac{\sin x}{x}\,\mathrm{d}x\right| < M + M = 2M,$$

i.e. the absolute value of integral (15) remains less than a definite positive number $2M$ for any positive a and b.

Before proving (13), we consider the simpler integral:

$$\frac{1}{\pi}\int_0^b \frac{\sin mx}{x}\,\mathrm{d}x.$$

If we change the variable in accordance with $t = mx$ and use (14), we get on indefinite increase of m:

$$\frac{1}{\pi}\int_0^b \frac{\sin mx}{x}\,\mathrm{d}x = \frac{1}{\pi}\int_0^{mb} \frac{\sin t}{t}\,\mathrm{d}t \to \frac{1}{\pi}\cdot\frac{\pi}{2} = \frac{1}{2},$$

and consequently:

$$\frac{1}{\pi}\int_0^b \varphi(+0)\frac{\sin mx}{x}\,\mathrm{d}x \to \frac{1}{2}\varphi(+0).$$

Hence it remains for us to show, in order to prove (13), that

$$\frac{1}{\pi}\int_0^b [\varphi(x)-\varphi(+0)]\frac{\sin mx}{x}\,\mathrm{d}x \to 0,$$

i.e. that the absolute value of the left-hand side is less than any positive ε for sufficiently large m. We split the interval of integration into two parts: $(0, \delta)$ and (δ, b), where δ is a small positive number which we fix later. We show that each of the two integrals:

$$\frac{1}{\pi}\int_0^\delta [\varphi(x)-\varphi(+0)]\frac{\sin mx}{x}\,\mathrm{d}x \quad\text{and}\quad \frac{1}{\pi}\int_\delta^b [\varphi(x)-\varphi(+0)]\frac{\sin mx}{x}\,\mathrm{d}x \tag{16}$$

has an absolute value less than $\varepsilon/2$ for sufficiently large m. Since $\varphi(x)$ has a finite number of discontinuities, we can take δ sufficiently small for the interval $(0, \delta)$ to contain no discontinuity, so that $\varphi(x \pm 0) = \varphi(x)$. On taking into account the fact that $\varphi(x)$ is monotonic, and applying the mean value theorem to the first of integrals (16), we get:

$$\frac{1}{\pi}\int_0^\delta [\varphi(x)-\varphi(+0)]\frac{\sin mx}{x}\,\mathrm{d}x = \frac{1}{\pi}[\varphi(\delta)-\varphi(+0)]\int_0^\delta \frac{\sin mx}{x}\,\mathrm{d}x,$$

and consequently:

$$\left|\frac{1}{\pi}\int_0^\delta [\varphi(x)-\varphi(+0)]\frac{\sin mx}{x}\,\mathrm{d}x\right| < \frac{1}{\pi}\left|\varphi(\delta)-\varphi(+0)\right|\cdot 2M.$$

We have $\varphi(\delta) - \varphi(+0) \to 0$ as $\delta \to 0$ by definition of the symbol $\varphi(+0)$, and we can therefore take δ close enough to zero for the right-hand side of the equation written above to be less than $\varepsilon/2$. The absolute value of the first of integrals (16) will now be less than $\varepsilon/2$ for any m. Having thus fixed the positive δ, we go back to the second of integrals (16). We write this in a new form by again applying the mean value theorem:

$$\frac{1}{\pi}[\varphi(\delta) - \varphi(+0)] \int_\delta^\xi \frac{\sin mx}{x}\,\mathrm{d}x + \frac{1}{\pi}[\varphi(b-0) - \varphi(+0)] \int_\xi^b \frac{\sin mx}{x}\,\mathrm{d}x. \quad (17)$$

The factors in front of the integrals are constants and it is sufficient for us to show that the integrals tend to zero on increase of m. We take say the first integral and change the variable in accordance with $t = mx$. We get:

$$\int_{m\delta}^{m\xi} \frac{\sin t}{t}\,\mathrm{d}t. \quad (18)$$

The limits $m\delta$ and $m\xi$ increase indefinitely on indefinite increase of m, since δ is a fixed positive number and ξ is not less than δ. But since

$$\int_0^\infty \frac{\sin t}{t}\,\mathrm{d}t$$

is convergent, integral (18) must tend to zero on indefinite increase of both its limits [82]. Similar arguments apply for the second integral in (17), so that the expression as a whole tends to zero, i.e. the second of integrals (16) tends to zero and its absolute value is less than $\varepsilon/2$ for sufficiently large m.

We have proved equation (13), and consequently all the statements of the lemma, on the assumption that $\varphi(z)$ is monotonic as well as that it satisfies Dirichlet conditions. It remains to show that (13) is still true when $\varphi(z)$ satisfies Dirichlet conditions only. These conditions imply that $(0, b)$ can be divided into a finite number of sub-intervals in each of which $\varphi(z)$ is monotonic. Let us suppose that $(0, b)$ can be divided into three such sub-intervals $(0, b_1)$, (b_1, b_2), (b_2, b). Integral (13) is now split into three:

$$\int_0^b \varphi(z)\frac{\sin mz}{z}\,\mathrm{d}z =$$

$$= \int_0^{b_1} \varphi(z)\frac{\sin mz}{z}\,\mathrm{d}z + \int_{b_1}^{b_2} \varphi(z)\frac{\sin mz}{z}\,\mathrm{d}z + \int_{b_2}^{b} \varphi(z)\frac{\sin mz}{z}\,\mathrm{d}z. \quad (19)$$

The lemma is applicable to each term on the right, since $\varphi(z)$ is monotonic in $(0, b_1)$, (b_1, b_2), and (b_2, b). Hence the first term tends to $\varphi(+0)/2$, whilst the remaining two tend to zero. It follows that integral (19) tends to $\varphi(+0)/2$, which it was required to prove.

It may be pointed out that the number m in Dirichlet integral (12) can increase indefinitely in any manner and is not obliged to take only integral values. The result obtained has its origin in the fact that the function $(1/z)\sin mz$ changes sign very often for large m and moreover, takes large values for z near zero.

153. Dirichlet's theorem. Dirichlet's theorem [143] can be proved without difficulty by using the lemma of the previous section. We have to show, by (3), that the expression

$$\frac{1}{\pi}\int_0^{\frac{\pi}{2}} f(x-2z)\,\frac{\sin(2n+1)z}{\sin z}\,dz+\frac{1}{\pi}\int_0^{\frac{\pi}{2}} f(x+2z)\,\frac{\sin(2n+1)z}{\sin z}\,dz \tag{20}$$

tends to $\{f(x-0)+f(x+0)\}/2$ on indefinite increase of n. Instead of (20), we take the expression:

$$\frac{1}{\pi}\int_0^{\frac{\pi}{2}} f(x-2z)\,\frac{\sin(2n+1)z}{z}\,dz+\frac{1}{\pi}\int_0^{\frac{\pi}{2}} f(x+2z)\,\frac{\sin(2n+1)z}{z}\,dz. \tag{21}$$

The upper limits are positive in both integrals, and the functions $f(x-2z)$ and $f(x+2z)$ satisfy Dirichlet conditions in the interval of integration. Furthermore, $m=2n+1\to\infty$, and by the lemma proved above, expression (21) tends to the limit $\{f(x-0)+f(x+0)\}/2$. It remains to show that the difference between expressions (20) and (21) tends to zero. For this, it is sufficient to show that the integrals

$$\frac{1}{\pi}\int_0^{\frac{\pi}{2}} f(x-2z)\left(\frac{1}{\sin z}-\frac{1}{z}\right)\sin(2n+1)z\,dz,$$

$$\frac{1}{\pi}\int_0^{\frac{\pi}{2}} f(x+2z)\left(\frac{1}{\sin z}-\frac{1}{z}\right)\sin(2n+1)z\,dz$$

tend to zero. We shall prove this for the first integral:

$$\frac{1}{\pi}\int_0^{\frac{\pi}{2}} f(x-2z)\left(\frac{1}{\sin z}-\frac{1}{z}\right)\sin(2n+1)z\,dz=\frac{1}{\pi}\int_0^{\frac{\pi}{2}}\varphi(z)\sin(2n+1)z\,dz, \tag{22}$$

where

$$\varphi(z)=f(x-2z)\left(\frac{1}{\sin z}-\frac{1}{z}\right).$$

The first factor $f(x - 2z)$ has a finite number of discontinuities of the first kind (or is continuous) in the interval of integration. The second factor

$$\frac{1}{\sin z} - \frac{1}{z} = \frac{z - \sin z}{z \sin z} = \frac{z - \left(\frac{z}{1!} - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots\right)}{z\left(\frac{z}{1!} + \frac{z^3}{2!} + \frac{z^5}{5!} - \dots\right)}$$

tends to zero as $z \to 0$ and has no discontinuities whatever in the interval $(0, \pi/2)$. Consequently the lemma of [150] is applicable to integral (22), and the integral tends to zero. The assertion of Dirichlet's theorem is thus proved.

We supplement the theorem with two further propositions which are stated without proof. The result obtained above reveals only that the Fourier series $S[f(x)]$ is convergent to the sum $f(x)$ at every point x of the interval; it says nothing about the nature of the convergence in the interval $(-\pi, \pi)$. This gap is filled by the propositions now stated:

1. *In every interval lying inside $(-\pi, \pi)$ in which $f(x)$ not only satisfies Dirichlet conditions but is also continuous, the series $S[f(x)]$ is uniformly convergent.*

2. *If $f(x)$ satisfies Dirichlet conditions and is continuous throughout $(-\pi, \pi)$, whilst moreover*

$$f(-\pi + 0) = f(\pi - 0),$$

the series $S[f(x)]$ is uniformly convergent for all x.

Dirichlet's theorem places relatively few restrictions on the function $f(x)$ to be expanded. Nevertheless, the theorem for expansion in Fourier series is not valid for any $f(x)$ and there even exist continuous functions which cannot be expanded in this way.

The reader will easily show that propositions similar to the above are valid for series expanded in cosines only or sines only for the case when the function is defined in $(0, \pi)$, with the following changes:

With the conditions of Dirichlet's theorem for $(0, \pi)$, the sum of the series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx; \quad a_k = \frac{2}{\pi} \int_0^{\pi} f(t) \cos kt \, dt \tag{23}$$

is equal to

$$\frac{f(x+0) + f(x-0)}{2} \qquad 0 < x < \pi \tag{24}$$

and to

$$f(+0) \quad \text{for} \quad x = 0; \qquad f(\pi - 0) \quad \text{for} \quad x = \pi;$$

whilst the sum of the series

$$\sum_{k=1}^{\infty} b_k \sin kx; \quad b_k = \frac{2}{\pi} \int_0^{\pi} f(t) \sin kt \, dt \tag{25}$$

is (24) for $0 < x < \pi$ and zero for $x = 0$ and $x = \pi$.

All these results are obtained very simply if $f(x)$ is continued in the neighbouring interval $(-\pi, 0)$ as was done in [145], the continuation being even in the case of series (23) and odd in the case of series (25).

154. Polynomial approximations to continuous functions. Our next task is to prove the closure equation (40) of [147]. The proof will be based on certain results in the theory of polynomial approximations. These results are important in themselves and their proof is based on the following theorem:

THEOREM I (Weierstrass's theorem). *If $f(x)$ is continuous in the closed interval $a \leqslant x \leqslant b$, a sequence of polynomials $P_1(x)$, $P_2(x)$, ... can be formed which tends uniformly* [I, 144] *to $f(x)$ throughout the closed interval (a, b).*

We notice first of all that the interval (a, b) can be reduced to $(0, 1)$ with the aid of the transformation $x' = (x-a)/(b-a)$, and polynomials in x become polynomials in x' and conversely. We can therefore assume that the interval (a, b) is $(0, 1)$. We start by proving two elementary algebraic identities. We write down the binomial formula:

$$\sum_{m=0}^{n} C_n^m u^m v^{n-m} = (u+v)^n. \tag{26}$$

On differentiating this with respect to u and multiplying by u, then carrying out the same process on the identity thus obtained, we arrive at the two new identities:

$$\left.\begin{aligned} &\sum_{m=0}^{n} m C_n^m u^m v^{n-m} = nu(u+v)^{n-1}, \\ &\sum_{m=0}^{n} m^2 C_n^m u^m v^{n-m} = nu(nu+v)(u+v)^{n-2}. \end{aligned}\right\} \tag{27}$$

On setting $u = x$ and $v = 1 - x$ in (26), we get:

$$1 = \sum_{m=0}^{n} C_n^m x^m (1-x)^{n-m}. \tag{28}$$

We now multiply (26) by $n^2 x^2$, the first of (27) by $(-2nx)$ and the second of (27) by unity, and add; this gives us, with $u = x$ and $v = 1 - x$:

$$\sum_{m=0}^{n} (m - nx)^2 C_n^m x^m (1-x)^{n-m} = nx(1-x).$$

It is easily shown [I, 60] that the right-hand side of this equation is positive in $(0, 1)$ and takes its greatest value for $x = 1/2$, whence it follows that

$$\sum_{m=0}^{n} (m - nx)^2 C_n^m x^m (1-x)^{n-m} \leqslant \frac{1}{4} n. \tag{29}$$

We now show that the polynomials

$$P_n(x) = \sum_{m=0}^{n} f\left(\frac{m}{n}\right) C_n^m x^m (1-x)^{n-m} \tag{30}$$

are uniformly convergent to $f(x)$ in $(0, 1)$. On multiplying both sides of (28) by $f(x)$ and subtracting (30) from the equation obtained, we can write:

$$f(x) - P_n(x) = \sum_{m=0}^{n} \left[f(x) - f\left(\frac{m}{n}\right)\right] C_n^m x^m (1-x)^{n-m}.$$

We have to show that, given a positive ε, there exists an N, not depending on x, such that:

$$\left|\sum_{m=0}^{n} \left[f(x) - f\left(\frac{m}{n}\right)\right] C_n^m x^m (1-x)^{n-m}\right| < \varepsilon \quad \text{for} \quad n > N.$$

Since the products $C_n^m x^m (1-x)^{n-m} \geqslant 0$ for $0 \leqslant x \leqslant 1$, we have

$$\left|\sum_{m=0}^{n} \left[f(x) - f\left(\frac{m}{n}\right)\right] C_n^m x^m (1-x)^{n-m}\right| \leqslant$$
$$\leqslant \sum_{m=0}^{n} \left|f(x) - f\left(\frac{m}{n}\right)\right| C_n^m x^m (1-x)^{n-m},$$

and it is sufficient to prove the inequality:

$$\sum_{m=0}^{n} \left|f(x) - f\left(\frac{m}{n}\right)\right| C_n^m x^m (1-x)^{n-m} < \varepsilon \quad \text{for} \quad n > N. \tag{31}$$

The function $f(x)$ is uniformly convergent in $(0, 1)$ [I, 35], i.e. there exists a δ such that $|f(x_1) - f(x_2)| < \varepsilon/2$ for $|x_1 - x_2| < \delta$. Let us fix x from $(0, 1)$. We split sum (31) in to two parts S_1 and S_2, and carry into the first sum the terms for which m satisfies the condition $|x - m/n| < \delta$. The first sum consists of positive terms and we can write, in view of the choice of δ:

$$S_1 < \sum_{(I)} \frac{\varepsilon}{2} C_n^m x^m (1-x)^{n-m},$$

where the (I) indicates that summation is over the m satisfying $|x - m/n| < \delta$. If the summation is over all m from 0 to n, the sum can only increase, i.e.

$$S_1 < \sum_{m=0}^{n} \frac{\varepsilon}{2} C_n^m x^m (1-x)^{n-m} = \frac{\varepsilon}{2} \sum_{m=0}^{n} C_n^m x^m (1-x)^{n-m}$$

so that by (28), $S_1 < \varepsilon/2$ for any n. We turn to the second sum

$$S_2 = \sum_{(II)} \left| f(x) - f\left(\frac{m}{n}\right) \right| C_n^m x^m (1-x)^{n-m},$$

where summation is over the m which satisfy $|x - m/n| \geqslant \delta$ or $|nx - m| \geqslant n\,\delta$, and evaluate it as follows. The function $f(x)$ is continuous in the closed interval (0, 1) and must therefore satisfy in the interval an inequality of the form $|f(x)| \leqslant M$, where M is a definite positive number [I, 35]; hence $|f(x) - f(m/n)| \leqslant |f(x)| + |f(m/n)| \leqslant 2M$. In addition, we multiply the terms of S_2 by the factors $(nx - m)^2/n^2\delta^2$, which are not less than unity. On taking outside the factors $2M$ and $1/n^2\delta^2$, which are independent of the variable of summation m, we get:

$$S_2 \leqslant \frac{2M}{n^2\,\delta^2} \sum_{(II)} (m - nx)^2 C_n^m x^m (1-x)^{n-m}.$$

All the terms are positive and the sum can only increase if the summation is over all m from $m = 0$ to $m = n$. On taking into account (29), we obtain:

$$S_2 \leqslant \frac{3M}{n^2\,\delta^2} \sum_{m=0}^{n} (m - nx)^2 C_n^m x^m (1-x)^{n-m} \leqslant \frac{M}{2n\delta^2}.$$

M and δ are both definite positive numbers and it is enough to take $M/2n\delta^2 < \varepsilon/2$, i.e. $n > M/\varepsilon\delta^2$, in order that S_2 may satisfy $S_2 < \varepsilon/2$. We have now obtained the $N = M/\varepsilon\delta^2$ which we required to find. For $n > N$, both S_1 and $S_2 < \varepsilon/2$ and inequality (31) is satisfied; Weierstrass's theorem is thus proved. As can easily be seen, the theorem may be stated as follows: *if $f(x)$ is continuous in the closed interval (a, b) and ε is any given positive number, there exists a polynomial $P(x)$ in x such that the inequality is satisfied throughout (a, b)*:

$$|f(x) - P(x)| < \varepsilon. \tag{32}$$

On the basis of Weierstrass's theorem, a similar theorem may be proved for periodic functions.

THEOREM II. *If $f(x)$ is a continuous periodic function of period 2π and ε is any given positive number, it is possible to find a trigonometric polynomial*

$$T(x) = c_0 + \sum_{k=1}^{m} (c_k \cos kx + d_k \sin kx), \tag{33}$$

such that for any x:

$$|f(x) - T(x)| < \varepsilon. \tag{34}$$

We notice first of all that, due to the periodicity, it is sufficient to prove inequality (34) in the basic interval $(-\pi, \pi)$. We start by supposing that $f(x)$ is an even function, and replace x by the new variable $t = \cos x$, i.e. $x = \arccos t$, the principal values of this function being taken, so that as t varies from 1 to (-1), $x = \arccos t$ varies continuously from 0 to π. The function $f(x) = f(\arccos t)$ will be continuous with respect to t in $(-1, 1)$ and in accordance with Weierstrass's theorem a polynomial $P(t)$ will exist such that

$$|f(\arccos t) - P(t)| < \varepsilon \quad (-1 \leqslant t \leqslant 1),$$

or on returning to the original variable,

$$|f(x) - P(\cos x)| < \varepsilon \qquad (0 \leqslant x \leqslant \pi).$$

Since $f(x)$ is even its value is unchanged on replacing x by $(-x)$ and similarly for $P(\cos x)$, since $\cos x$ is even; hence the inequality written is also valid for $-\pi \leqslant x \leqslant 0$, i.e. it is valid throughout the fundamental interval. We know from [I, 176] that positive integral powers of $\sin x$ and $\cos x$ can be expressed linearly in terms of sines and cosines of multiples of the angle, i.e. $P(\cos x)$ can be written in the form (33); thus the theorem is proved.

We now take any continuous periodic function $f(x)$. If we put

$$\varphi(x) = \frac{1}{2}[f(x) + f(-x)]; \quad \psi(x) = \frac{1}{2}[f(x) - f(-x)], \tag{35}$$

$f(x)$ is equal to the sum of $\varphi(x)$ and $\psi(x)$, where $\varphi(x)$ and $\psi(x)$ are respectively even and odd periodic functions. By what has been proved, there exists for a given ε a polynomial $P(t)$ such that $|\varphi(x) - P(\cos x)| < \varepsilon/2$. If we can show that a polynomial $Q(t)$ exists such that

$$|\psi(x) - \sin x\, Q(\cos x)| < \frac{\varepsilon}{2} \, (-\pi \leqslant x \leqslant \pi), \tag{36}$$

the trigonometric polynomial

$$T(x) = P(\cos x) + \sin x\, Q(\cos x)$$

will satisfy condition (34). We introduce as before a new variable $t = \cos x$ and consider $\psi(x) = \psi(\arccos t)$ in the interval $-1 \leqslant t \leqslant 1$. In the manner of all continuous, odd periodic functions, $\psi(x)$ vanishes for $x = 0$ and $x = \pi$, and consequently $\psi(\arccos t)$ vanishes at the

ends of the interval, i.e. at $t = \pm 1$. It follows from (30) that if $f(x)$ vanishes at the ends of the interval (0, 1), i.e. $f(0) = f(1) = 0$, the polynomial $P_n(x)$ has the same property. We can transform the interval (0, 1) to (−1, 1) by using the transformation $t = 2x - 1$, which enables us to say that a polynomial $R(t)$ exists, equal to zero at $t = \pm 1$, and such that

$$|\psi(\operatorname{arc}\cos t) - R(t)| < \frac{\varepsilon}{4} \quad \text{for} \quad -1 \leqslant t \leqslant 1.$$

We can now write $R(t) = (1 - t^2)R_1(t)$, where $R_1(t)$ is also a polynomial, and the above inequality can be written in the new form:

$$|\psi(x) - \sin^2 x R_1(\cos x)| \leqslant \frac{\varepsilon}{4} \quad \text{for} \quad 0 \leqslant x \leqslant \pi. \tag{37}$$

As regards the function $\sin x\, R_1(\cos x) = \sqrt{1 - t^2}\, R_1(t)$, continuous in (−1, 1), we can find a polynomial $Q(t)$ such that

$$\left|\sqrt{1 - t^2}\, R_1(t) - Q(t)\right| < \frac{\varepsilon}{4} \quad \text{for} \quad -1 \leqslant t \leqslant 1,$$

i.e.

$$|\sin x R_1(\cos x) - Q(\cos x)| < \frac{\varepsilon}{4} \quad \text{for} \quad 0 \leqslant x \leqslant \pi,$$

and all the more,

$$|\sin^2 x R_1(x) - \sin x\, Q(x)| < \frac{\varepsilon}{4}, \tag{37_1}$$

since $|\sin x| \leqslant 1$. It follows from (37) and (37_1) that

$$|\psi(x) - \sin x\, Q(\cos x)| \leqslant |\psi(x) - \sin^2 x\, R_1(\cos x)| +$$

$$+ |\sin^2 x\, R_1(\cos x) - \sin x\, Q(\cos x)| \leqslant \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2},$$

i.e. we have proved inequality (36) for the interval $(0, \pi)$. But $\psi(x)$ and $\sin x\, Q(\cos x)$ are both odd, so that the inequality likewise applies throughout $(-\pi, \pi)$.

The above proofs of Theorems I and II are due to Prof. S. N. Bernshtein.

155. The closure equation. The closure equation of [147] for systems of trigonometric functions follows fairly easily from the theorem just proved. We start by supposing that $f(x)$ is a continuous function given in the interval $(-\pi, \pi)$ and that $f(-\pi) = f(\pi)$.

We obtain a continuous periodic function by periodic continuation of $f(x)$ outside the interval. There will exist, for a given ε, a trigonometric polynomial $T(x)$ which satisfies inequality (34).

It follows from this inequality that

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)-T(x)]^2\,dx<\varepsilon^2. \tag{38}$$

Let n be the order of the trigonometric polynomial, i.e. the value of m in expression (33). Now on the other hand, for any choice of trigonometric polynomial of order not higher than n, integral (38) has a least value ε_n^2, obtained when we take as the polynomial the sum of the first $(2n+1)$ terms of the Fourier series for $f(x)$. It follows from this that $\varepsilon_n < \varepsilon$, and since we can take as small a positive ε as we wish, we can say that ε_n, which does not increase with increasing n, must tend to zero as $n\to\infty$; and as we know from [147], this is equivalente to the closure equation for $f(x)$.

We now take the more general case when $f(x)$ is continuous in $(-\pi, \pi)$ but $f(-\pi)$ and $f(\pi)$ are distinct. As usual, a positive number M exists such that $|f(x)|\leqslant M$ for $-\pi\leqslant x\leqslant\pi$. Let a positive number η be arbitrarily assigned and let the positive δ satisfy the inequalities:

$$\delta<\frac{\pi\eta}{8M^2}\ ;\ \ \delta<\pi. \tag{39}$$

We construct a new function $f_1(x)$ in accordance with the following rule. We take $f_1(x)$ equal to $f(x)$ in $(-\pi, \pi-\delta)$, whilst we take the graph of $f_1(x)$ in $(\pi-\delta, \pi)$ as the straight line joining the points $x=\pi-\delta$, $y=f(\pi-\delta)$ and $x=\pi$, $y=f(-\pi)$ (Fig. 126). Now $f_1(x)$ is continuous in $(-\pi, \pi)$ and has the same value $f(-\pi)$ for $x = \pm\pi$, whilst evidently, as for $f(x)$, $|f_1(x)|\leqslant M$.

By what we have proved above, for any given positive η a trigonometric polynomial can be found such that

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f_1(x)-T(x)]^2\,dx<\frac{\eta}{4}\,. \tag{40}$$

Since $f(x)=f_1(x)$ in $(-\pi, \pi-\delta)$, we have:

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)-f_1(x)]^2\,dx=\frac{1}{2\pi}\int_{\pi-\delta}^{\pi}[f(x)-f_1(x)]^2\,dx.$$

Hence we can write, since $|f(x) - f_1(x)| \leqslant |f(x)| + |f_1(x)| \leqslant 2M$:

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)-f_1(x)]^2\,dx \leqslant \frac{2M^2}{\pi}\int_{\pi-\delta}^{\pi}dx = \frac{2M^2\delta}{\pi},$$

or, by (39):

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)-f_1(x)]^2\,dx < \frac{\eta}{4}. \tag{41}$$

We now consider:

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)-T(x)]^2\,dx = \frac{1}{2\pi}\int_{-\pi}^{+\pi}\{[f(x)-f_1(x)]+[f_1(x)-T(x)]\}^2\,dx.$$

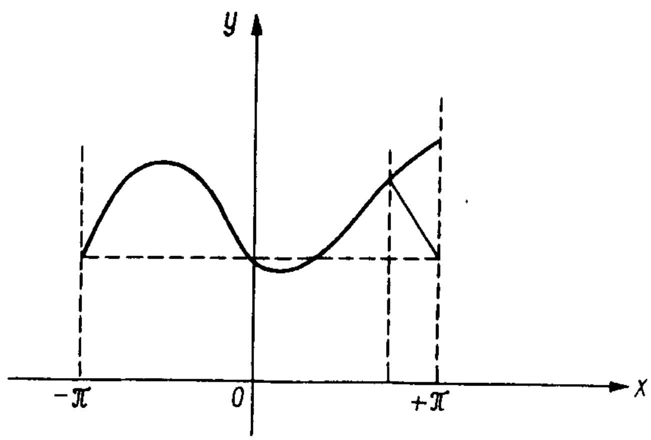

FIG. 126

On taking into account the obvious inequality $(a + b)^2 \leqslant 2(a^2 + b^2)$, we can write:

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)-T(x)]^2\,dx \leqslant$$
$$\leqslant \frac{1}{\pi}\int_{-\pi}^{+\pi}[f(x)-f_1(x)]^2\,dx + \frac{1}{\pi}\int_{-\pi}^{+\pi}[f_1(x)-T(x)]^2\,dx,$$

and it follows from this, by (40) and (41), that

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi}[f(x)-T(x)]^2\,dx < \eta.$$

If we let n denote the order of the trigonometric polynomial $T(x)$ and argue as above, we obtain from this $\varepsilon_n^2 \leqslant \eta$, and in view of the arbi-

trary smallness of η we have $\varepsilon_n \to 0$ as $n \to \infty$, i.e. the closure equation is also valid for $f(x)$ with the properties stated above. It can be shown in exactly the same way that this equation likewise holds for $f(x)$ bounded in $(-\pi, \pi)$ and having a finite number of discontinuities. If all the discontinuities are of the first kind, there is no need to stipulate boundedness of the function. We can carry out the proof by isolating the points of discontinuity with narrow intervals then forming a new function $f_1(x)$, continuous in $(-\pi, \pi)$ and coinciding with $f(x)$ outside these intervals, whilst having a linear graph inside the intervals. By the above, a trigonometric polynomial $T(x)$ which satisfies inequality (40) can be formed for $f_1(x)$, whilst the isolating intervals can be chosen narrow enough for inequality (41) to be satisfied. The rest of the proof is as above. We have thus proved the closure equation for all functions having a finite number of discontinuities of the first kind (or which are continuous). It may be mentioned that the equation holds for a much wider class of functions.

156. Properties of closed systems of functions. We now consider some consequences of the closure equation, and instead of confining ourselves to systems of trigonometric functions, refer the discussion to any system of orthogonal and normalized functions

$$\psi_1(x), \psi_2(x), \psi_3(x), \ldots, \psi_n(x), \ldots \tag{42}$$

in the interval (a, b). We suppose that, with respect to this system, the closure equation is valid for any function with a finite number of discontinuities of the first kind. We shall only refer to functions of this type in future. We bring in the generalized Fourier coefficients of the function $f(x)$:

$$c_k = \int_a^b f(x)\, \psi_k(x)\, \mathrm{d}x.$$

The closure equation has the form:

$$\int_a^b [f(x)]^2\, \mathrm{d}x = \sum_{k=1}^{\infty} c_k^2. \tag{43}$$

We now turn our attention to some important consequences of this formula.

1. *If $f(x)$ and $\varphi(x)$ are any two functions, with Fourier coefficients c_k and d_k:*

$$c_k = \int_a^b f(x)\, \varphi_k(x)\, \mathrm{d}x; \qquad d_k = \int_a^b \varphi(x)\, \psi_k(x)\, \mathrm{d}x, \tag{44}$$

we have

$$\int_a^b f(x)\, \varphi(x)\, \mathrm{d}x = \sum_{k=1}^{\infty} c_k d_k, \tag{45}$$

the series on the right being absolutely convergent.

In fact, on replacing $f(x)$ in (43) by $f(x) + \xi\varphi(x)$, where ξ is an arbitrary constant parameter, we obtain:

$$\int_a^b [f(x) + \xi\varphi(x)]^2 \, dx = \sum_{k=1}^{\infty} [\int_a^b f(x) + \xi\varphi(x)\, \psi_k(x)\, dx]^2,$$

or, by (44):

$$\int_a^b [f(x)]^2 \, dx + 2\xi \int_a^b f(x)\,\varphi(x)\, dx + \xi^2 \int_a^b [\varphi(x)]^2 \, dx =$$

$$= \sum_{k=1}^{\infty} c_k^2 + 2\xi \sum_{k=1}^{\infty} c_k d_k + \xi^2 \sum_{k=1}^{\infty} d_k^2.$$

Comparison of the coefficients of like powers of ξ gives us (45).

The absolute convergence of the series appearing in (45) is obvious from the fact that

$$|c_k d_k| \leqslant \frac{1}{2}(c_k^2 + d_k^2).$$

since we know that the series $\sum_{k=1}^{\infty} (c_k^2 + d_k^2)$ is convergent.

2. *If a function $\varphi(x)$, dependent on certain parameters, satisfies for all values of its parameters*

$$\int_a^b [\varphi(x)]^2 \, dx < M,$$

where M is an independent constant, the series

$$\sum_{k=1}^{\infty} c_k d_k \tag{46}$$

is uniformly convergent with respect to the parameters.

The proof is based on a simple, yet at the same time extremely important inequality: *whatever the real constants*

$$\alpha_1, \alpha_2, \ldots, \alpha_m;\ \beta_1, \beta_2, \ldots, \beta_m,$$

we always have

$$\left(\sum_{k=1}^{m} \alpha_k \beta_k\right)^2 \leqslant \sum_{k=1}^{m} \alpha_k^2 \cdot \sum_{k=1}^{m} \beta_k^2, \tag{47}$$

the sign of equality being obtained only when all the α_i and β_i are in the same ratio:

$$\frac{\beta_1}{\alpha_1} = \frac{\beta_2}{\alpha_2} = \ldots = \frac{\beta_m}{\alpha_m}.$$

In fact, let ξ be any real number. We form the sum

$$S_m = \sum_{k=1}^{m} (\xi\alpha_k - \beta_k)^2, \tag{48}$$

which is clearly non-negative. The sum can only be zero in the case when

$$\xi a_k - \beta_k = 0, \qquad k = 1, 2, \ldots, m,$$

i.e.

$$\frac{\beta_1}{a_1} = \frac{\beta_2}{a_2} = \ldots = \frac{\beta_m}{a_m} = \xi,$$

and evidently in this case:

$$\left(\sum_{k=1}^{m} a_k \beta_k\right)^2 = \sum_{k=1}^{m} a_k^2 \cdot \sum_{k=1}^{m} \beta_k^2.$$

Generally speaking, on removing the brackets in (48), we can write an equation of the form

$$S_m = A\xi^2 - 2B\xi + C,$$

where

$$A = \sum_{k=1}^{m} a_k^2; \quad B = \sum_{k=1}^{m} a_k \beta_k; \quad C = \sum_{k=1}^{m} \beta_k^2,$$

Since S_m must always remain positive, we know from elementary algebra that $B^2 - AC < 0$, i.e. we must have $B^2 < AC$, which gives inequality (47).

We return to our proposition 2. We form the sum:

$$\sum_{k=n+1}^{n+p} c_k d_k;$$

we have in accordance with (47):

$$\left|\sum_{k=n+1}^{n+p} c_k d_k\right| < \sqrt{\sum_{k=n+1}^{n+p} c_k^2} \cdot \sqrt{\sum_{k=n+1}^{n+p} d_k^2}.$$

On the other hand, if we replace $f(x)$ and c_k in (43) by $\varphi(x)$ and d_k respectively we are enabled to write:

$$\sum_{k=n+1}^{n+p} d_k^2 < \sum_{k=1}^{\infty} d_k^2 = \int_a^b [\varphi(x)]^2 \, dx < M.$$

The terms of the series $\sum_{k=1}^{\infty} c_k^2$ are independent of the parameters by hypothesis, so that for any previously assigned small positive ε we can find an N, independent of the parameters, such that for all $n > N$ and all $p > 0$ we have the inequality:

$$\sum_{k=n+1}^{n+p} c_k^2 < \frac{\varepsilon^2}{M}.$$

With this, we get:

$$\left|\sum_{k=n+1}^{n+p} c_k d_k\right| < \varepsilon \quad \text{for} \quad n > N,$$

from which follows the uniform convergence of series (46).

3. *If x_1 and x_2 are any two values from the interval (a, b), we have:*

$$\int_{x_1}^{x_2} f(x)\,\mathrm{d}x = \sum_{k=1}^{\infty} c_k \int_{x_1}^{x_2} \psi_k(x)\,\mathrm{d}x, \tag{49}$$

the series on the right being uniformly convergent for all x_1, x_2 of (a, b).

If we knew that $f(x)$ had the Fourier expansion

$$f(x) = \sum_{k=1}^{\infty} c_k \psi_k(x) \tag{50}$$

and that this series was uniformly convergent, (49) would be obvious [I, 146].

The remarkable fact is that the formula is always valid, even if series (50) is not convergent, i.e. (50) can be integrated term by term just as though it were uniformly convergent and had the sum $f(x)$.

To prove (49), we put in expression (45):

$$\varphi(x) = \begin{cases} 1 \text{ if } x_1 \leqslant x \leqslant x_2 \\ 0 \text{ if } a \leqslant x < x_1 \text{ or } x_2 < x \leqslant b. \end{cases}$$

The quantities x_1 and x_2 appear here as the parameters on which $\varphi(x)$ depends. The number M mentioned in proposition 2 exists, since

$$\int_a^b [\varphi(x)]^2\,\mathrm{d}x \leqslant \int_a^b \mathrm{d}x = b - a.$$

We have furthermore, since $\varphi(x)$ is zero outside the interval (x_1, x_2):

$$d_k = \int_a^b \varphi(x)\,\psi_k(x)\,\mathrm{d}x = \int_{x_1}^{x_2} \psi_k(x)\,\mathrm{d}x$$

and, by (45):

$$\int_a^b f(x)\,\varphi(x)\,\mathrm{d}x = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x = \sum_{k=1}^{\infty} c_k d_k = \sum_{k=1}^{\infty} c_k \int_{x_1}^{x_2} \psi_k(x)\,\mathrm{d}x,$$

which it was required to prove.

Remark 1. On applying this to ordinary Fourier series, it can be shown that they can be integrated term by term as though they were uniformly convergent and had the sum $f(x)$ not only for intervals inside $(-\pi, \pi)$ but for any interval in general. Here, the function $f(x)$ must be continued periodically outside $(-\pi, \pi)$ as indicated in [143].

Remark 2. Inequality (47) is applicable to integrals as well as sums, in which case it has the form (Buniakowski's inequality):

$$\left[\int_a^b f_1(x)\,f_2(x)\,\mathrm{d}x\right]^2 \leqslant \int_a^b f_1^2(x)\,\mathrm{d}x \cdot \int_a^b f_2^2(x)\,\mathrm{d}x. \tag{51}$$

We see this by considering

$$\int_a^b [f_1(x) + \xi f_2(x)]^2\,\mathrm{d}x = \int_a^b f_1^2(x)\,\mathrm{d}x + 2\xi \int_a^b f_1(x)\,f_2(x)\,\mathrm{d}x + \xi^2 \int_a^b f_2^2(x)\,\mathrm{d}x,$$

where ξ is any real number. It follows at once from the form of the left-hand side that this cannot be negative for real ξ. But if $A + 2B\xi + C\xi^2$ is non-negative for real ξ, we must have $B^2 - AC \leqslant 0$, i.e. $B^2 \leqslant AC$. Application of this to the above gives us inequality (51).

157. The character of the convergence of Fourier series. The series obtained by us in [144] suffer from the defect that they converge badly. Some of them are not absolutely and uniformly convergent, for instance, series (10) [144] becomes for $x = \pi/2$:

$$2\left(\frac{1}{1} - \frac{1}{3} + \frac{1}{5} - \dots\right)$$

which is not absolutely convergent; in addition, (10) cannot be uniformly convergent since it represents a discontinuous function [I, 146]. The series representing the discontinuous function having the values c_1 and c_2 has the same defect. There is a relationship between the nature of the expanded function, its continuity, and its Fourier series. We shall consider the relationship in detail here. We assume once and for all that the $f(x)$ and its successive derivatives, to be discussed later, are functions satisfying Dirichlet conditions and periodically continued outside $(-\pi, \pi)$. Let

$$x_1^{(0)},\ x_2^{(0)},\ \dots,\ x_{\tau_0-1}^{(0)}$$

denote the discontinuities of $f(x)$ inside $(-\pi, \pi)$, and let

$$x_1',\ x_2',\ \dots,\ x_{\tau_1-1}'$$

denote the discontinuities of its derivative $f'(x)$ inside $(-\pi, \pi)$, whilst in general

$$x_1^{(k)},\ x_2^{(k)},\ \dots,\ x_{\tau_k-1}^{(k)}$$

are the discontinuities of the derivative $f^{(k)}(x)$. It will be necessary to add to the discontinuities the ends of the interval $(-\pi, \pi)$, if the limits:

$$f(\mp\pi \pm 0),\quad f'(\mp\pi \pm 0),\ \dots,\ f^{(k)}(\mp\pi \pm 0)$$

are not the same.

For the sake of symmetry, we write $x_0^{(0)} = -\pi$ and $x_{\tau_0}^{(0)} = \pi$, and similarly for the derivatives. The above condition for the derivatives amounts to the existence of a continuous $f^{(k)}(x)$ inside any interval $(x_s^{(k)}, x_{s+1}^{(k)})$ $(s = 0, 1, \dots, \tau_{k-1})$. By the Dirichlet conditions, the derivatives will tend to definite limits at the ends of the interval.

We now transform the expression for the Fourier coefficients of $f(x)$. We start with the coefficient

$$a_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\cos nx\,dx.$$

We sub-divide the interval of integration $(-\pi, \pi)$ into

$$(-\pi,\ x_1^{(0)}),\ (x_1^{(0)},\ x_2^{(0)}),\dots,\ (x_{\tau_0-1}^{(0)},\ \pi),$$

so that $f(x)$ is continuous in each sub-interval. Integration by parts gives us:

$$\int f(x) \cos nx \, dx = \frac{\sin nx}{n} f(x) - \frac{1}{n} \int f'(x) \sin nx \, dx,$$

Since, on the other hand:

$$\int_{x_{i-1}^{(0)}}^{x_i^{(0)}} f(x) \cos nx \, dx = \lim_{\varepsilon', \varepsilon'' \to 0} \int_{x_{i-1}^{(0)}+\varepsilon'}^{x_i^{(0)}-\varepsilon''} f(x) \cos nx \, dx =$$

$$= \lim_{\varepsilon', \varepsilon'' \to 0} \frac{\sin nx}{n} f(x) \Big|_{x=x_{i-1}^{(0)}+\varepsilon'}^{x=x_i^{(0)}-\varepsilon''} - \frac{1}{n} \int_{x_{i-1}^{(0)}}^{x_i^{(0)}} f'(x) \sin nx \, dx,$$

we obtain in view of the continuity of $\sin nx$:

$$\int_{x_{i-1}^{(0)}}^{x_i^{(0)}} f(x) \cos nx \, dx = \frac{\sin nx_i^{(0)}}{n} f(x_i^{(0)} - 0) - \frac{\sin nx_{i-1}^{(0)}}{n} f(x_{i-1}^{(0)} + 0) -$$

$$- \frac{1}{n} \int_{x_{i-1}^{(0)}}^{x_i^{(0)}} f'(x) \sin nx \, dx.$$

We finally have, on summing over i from 1 to τ_0:

$$a_n = - \frac{1}{\pi n} \{\sin nx_1^{(0)} [f(x_1^{(0)} + 0) - f(x_1^{(0)} - 0)] + \ldots +$$

$$+ \sin nx_{\tau_0}^{(0)} [f(x_{\tau_0}^{(0)} + 0) - f(x_{\tau_0}^{(0)} - 0)]\} - \frac{1}{n\pi} \int_{-\pi}^{+\pi} f'(x) \sin nx \, dx,$$

where $x_0^{(0)} = -\pi$, $x_{\tau_0}^{(0)} = +\pi$, and by the periodicity of $f(x)$, $f(x_{\tau_0}^{(0)} + 0) = f(x_0^{(0)} + 0)$. In the present case $\sin nx_{\tau_0}^{(0)} = 0$, but we preserve the term for the sake of the symmetry of later expressions.

For brevity, we denote the jumps of $f(x)$ at its points of discontinuity $x_1^{(0)}$, $x_2^{(0)}, \ldots, x_{\tau_0}^{(0)}$, as respectively:

$$\delta_1^{(0)} = f(x_1^{(0)} + 0) - f(x_1^{(0)} - 0); \ldots, \; \delta_{\tau_0}^{(0)} = f(x_{\tau_0}^{(0)} + 0) - f(x_{\tau}^{(0)} - 0).$$

The formula above can now be written in the form:

$$a_n = - \frac{1}{\pi n} \sum_{i=1}^{\tau_0} \delta_i^{(0)} \sin nx_i^{(0)} - \frac{b_n'}{n}, \tag{52}$$

where a_n' and b_n' denote the Fourier coefficients of the derivative $f'(x)$. Similarly, on departing from the expression:

$$\int f(x) \sin nx \, dx = - \frac{\cos nx}{n} f(x) + \int f'(x) \cos nx \, dx,$$

we obtain:

$$b_n = \frac{1}{\pi n} \sum_{i=1}^{\tau_0} \delta_i^{(0)} \cos nx_i^{(0)} + \frac{a_n'}{n}. \tag{53}$$

Expressions (52) and (53) are important in themselves, since they show that *if a periodic function $f(x)$ has jumps, its Fourier coefficients are of order $1/n$ as $n \to \infty$, the main parts of the coefficients a_n and b_n being respectively equal to:*

$$-\frac{1}{\pi n} \sum_{i=1}^{\tau_0} \delta_i^{(0)} \sin nx_i^{(0)}; \quad \frac{1}{\pi n} \sum_{k=1}^{\tau_0} \delta_i^{(0)} \cos nx_i^{(0)}, \tag{54}$$

whilst their remainders are of higher order than $1/n$.

The remainders are in fact:

$$-\frac{b_n'}{n}, \quad \frac{a_n'}{n};$$

and since a_n', b_n' are the Fourier coefficients of $f'(x)$, they tend to zero as $n \to \infty$, i.e. they become infinitesimals as $n \to \infty$. There is another reason for the importance of (52) and (53); by using them, we can *distinguish in the Fourier coefficients a_n and b_n, which tend to zero as $n \to \infty$, the components of different orders of smallness with respect to $1/n$.*

To do this, we let $a_n^{(k)}$, $b_n^{(k)}$ denote the Fourier coefficients of the kth order derivative $f^{(k)}(x)$, whilst $\delta_1^{(k)}, \ldots, \delta_{\tau_k}^{(k)}$ are their jumps at the points $x_1^{(k)}, \ldots, x_{\tau_k}^{(k)} = \pi$:

$$\delta_1^{(k)} = f^{(k)}(x_1^{(k)} + 0) - f^{(k)}(x_1^{(k)} - 0); \ldots; \quad \delta_{\tau_k}^{(k)} = f^{(k)}(\pi + 0) - f^{(k)}(\pi - 0).$$

We apply (52) and (53) to coefficients a_n', b_n', which merely requires the eplacing of $f(x)$ by $f'(x)$, $\delta_i^{(0)}$ by $\delta_i^{(1)}$, $x_i^{(0)}$ by $x_i^{(1)}$, and τ_0 by τ_1; we get:

$$\left.\begin{aligned} a_n' &= -\frac{1}{\pi n} \sum_{i=1}^{\tau_1} \delta_i^{(1)} \sin nx_i^{(1)} - \frac{b_n''}{n}, \\ b_n' &= \frac{1}{\pi n} \sum_{i=1}^{\tau_1} \delta_i^{(1)} \cos nx_i^{(1)} + \frac{a_n''}{n}, \end{aligned}\right\} \tag{55}$$

where a_n'', b_n'' are the Fourier coefficients of $f''(x)$.

Repetition of the above argument gives us:

$$\left.\begin{aligned} a_n'' &= -\frac{1}{\pi n} \sum_{i=1}^{\tau_2} \delta_i^{(2)} \sin nx_i^{(2)} - \frac{b_n'''}{n}, \\ b_n'' &= \frac{1}{\pi n} \sum_{i=1}^{\tau_2} \delta_i^{(2)} \cos nx_i^{(2)} + \frac{a_n'''}{n}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}\right\} \tag{56}$$

If we write for brevity:

$$A_k = \frac{1}{\pi} \sum_{i=1}^{\tau_k} \delta_i^{(k)} \sin nx_i^{(k)}; \qquad B_k = \frac{1}{\pi} \sum_{i=1}^{\tau_k} \delta_i^{(k)} \cos nx_i^{(k)} \quad (k = 0, 1, 2, \ldots),$$

we have from the above expressions:

$$\left.\begin{aligned} a_n &= -\frac{A_0}{n} - \frac{B_1}{n^2} + \frac{A_2}{n^3} + \frac{B_3}{n^4} - \ldots + \frac{\varrho_k'}{n^k}, \\ b_n &= \frac{B_0}{n} - \frac{A_1}{n^2} - \frac{B_2}{n^3} + \frac{A_3}{n^4} + \ldots + \frac{\varrho_k''}{n^k}, \end{aligned}\right\} \quad (57)$$

where ϱ_k', ϱ_k'' have different expressions depending on the form of the number k, as given in the following table:

k	$4m$	$4m+1$	$4m+2$	$4m+3$
ϱ_k'	$a_n^{(k)}$	$-b_n^{(k)}$	$-a_n^{(k)}$	$b_n^{(k)}$
ϱ_k''	$b_n^{(k)}$	$a_n^{(k)}$	$-b_n^{(k)}$	$-a_n^{(k)}$

Here, $a_n^{(k)}$ and $b_n^{(k)}$ are the Fourier coefficients of $f^{(k)}(x)$.

It is clear from the expressions for A_k and B_k that these are dependent on n; however, n only appears under the sign of the trigonometric function, so that with fixed s, A_s and B_s remain bounded. The coefficients of the trigonometric functions in A_s and B_s consist of the jumps of the derivative $f^{(s)}(x)$. If there are no jumps, $A_s = B_s = 0$. On the other hand, if $f^{(k)}(x)$ satisfies Dirichlet conditions, the factors ϱ_k' and ϱ_k'', which coincide except for the sign with one of the Fourier coefficients of $f^{(k)}(x)$, will be of order not lower than $1/n$ for large n, since we saw from [151] that the Fourier coefficients of a function that satisfies Dirichlet conditions are of order not lower than $1/n$. We thus obtain the following theorem:

If a continuous periodic function $f(x)$ has continuous derivatives up to and including the $(k-1)$-th order, whilst the k-th order derivative satisfies Dirichlet conditions, the Fourier coefficients a_n and b_n of $f(x)$ are of order not lower than $1/n^{k+1}$, i.e.

$$|a_n| \leqslant \frac{M}{n^{k+1}}; \qquad |b_n| \leqslant \frac{M}{n^{k+1}},$$

where M is a positive number.

It may be remarked that, for $k \geqslant 1$, the Fourier series for $f(x)$ is uniformly convergent. It follows in fact from the theorem proved that in this case a_n and b_n satisfy the inequalities

$$|a_n| < \frac{M}{n^2}; \quad |b_n| < \frac{M}{n^2},$$

whilst we can write for the general term of the series:

$$|a_n \cos nx + b_n \sin nx| < \frac{2M}{n^2},$$

whence follows the absolute and uniform convergence of the series, since $\sum\limits_{n=1}^{\infty} 1/n^2$ is convergent [I, 122].

Formula (57) remains valid for Fourier expansions in the interval $(-l, l)$. We merely have to put

$$\left.\begin{aligned} A_k &= \left(\frac{l}{\pi}\right)^k \frac{1}{\pi} \sum_{i=1}^{\tau_k} \delta_i^{(k)} \sin \frac{n\pi x_i^{(k)}}{l}, \\ B_k &= \left(\frac{l}{\pi}\right)^k \frac{1}{\pi} \sum_{i=1}^{\tau_k} \delta_i^{(k)} \cos \frac{n\pi x_i^{(k)}}{l}, \end{aligned} \quad (k = 0, 1, 2, \ldots) \right\} \tag{58}$$

whilst the expressions written in the table for ϱ_k', ϱ_k'' have to be multiplied by $(l/\pi)^k$, whilst now

$$\delta_{\tau_k}^{(k)} = f^{(k)}(l+0) - f^{(k)}(l-0) = f^{(k)}(-l+0) - f^{(k)}(-l-0) \ldots \tag{59}$$

158. Improving the convergence of Fourier series. As we saw above, the presence of terms of order $1/n$ in the expressions for the Fourier coefficients a_n and b_n of $f(x)$, which cause the slowness of convergence, are due to the jumps in $f(x)$. However many derivatives the function may have in $(-\pi, \pi)$, it only needs one jump at the end of the interval, or to be precise, distinctness of the limits $f(\mp\pi \pm 0)$, for its Fourier series to be unsuitable for practical purposes. Furthermore, the important thing in applications is often not the investigation of the expansion of $f(x)$ itself but that of its first, second, or even third order derivatives. If the coefficients of $f(x)$ itself are of order $1/n^{k+1}$, differentiation of the series brings the coefficients to the order $1/n^k$, as is clear from the equations:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

$$f'(x) = \sum_{n=1}^{\infty} n(b_n \cos nx - a_n \sin nx),$$

$$f''(x) = \sum_{n=1}^{\infty} n^2 (-a_n \cos nx - b_n \sin nx).$$

Conversely, the order of the coefficients is increased by unity with each integration, since

$$\int \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)\,\mathrm{d}x = C + \sum_{n=1}^{\infty} \frac{-b_n \cos nx + a_n \sin nx}{n},$$

where C is an arbitrary constant.

The convergence of a Fourier series thus deteriorates on differentiation; if say the coefficients of $f(x)$ are of order $1/n^2$, as is the case for $f(x)$ continuous and periodic, where $f'(x)$ may have discontinuities, the series obtained for $f'(x)$ by term-by-term differentiation will have coefficients of order $1/n$, whilst the series for $f''(x)$ becomes quite meaningless, inasmuch as its coefficients do not even tend to zero. It may happen that the Fourier expansion of $f(x)$ is entirely unsuitable for evaluation of its derivatives for any value of x, this being the case for an $f(x)$ which, whilst possessing derivatives of any order at all other points, lacks a derivative merely at *a single point*.

The problem therefore arises of *improving the convergence of Fourier series*, i.e. of transforming them to new series in such a way that the new coefficients have a high enough order of smallness for the deterioration in convergence following on differentiation not to prevent the evaluation of derivatives. For instance, if we want to be free to evaluate the derivatives up to and including the third order by term-by-term differentiation, we want the original series to have coefficients of order not lower than $1/n^5$, since we then get coefficients of order $1/n^2$ in the series for the third order derivative, this latter series being uniformly convergent and suitable for practical computations.

We can improve the convergence of the Fourier series for $f(x)$ as follows. Let there be terms of order $1/n$ in formulae (57), i.e. $f(x)$ has jumps $\delta_l^{(0)}$.

A simple auxiliary function $\varphi_0(x)$ can always be found, having the same jumps as $f(x)$. The difference:

$$f_1(x) = f(x) - \varphi_0(x)$$

now has no jumps, and its Fourier expansion $S(f_1)$ has coefficients of order at least $1/n^2$. The simplest choice for $\varphi_0(x)$ is the function whose graph is a step-line i.e. consists of a number of lines either parallel to OX, or with say constant slope m_0; in the first case we have:

$$\varphi_0'(x) = 0, \quad \text{i.e.} \quad f_1'(x) = f'(x),$$

whilst in the second:

$$f_1'(x) - f'(x) = -m_0,$$

so that $f_1'(x)$ has the same jumps as $f'(x)$.

We take $\varphi_0(x)$ as having been defined in some such way. We have:

$$f(x) = \varphi_0(x) + f_1(x),$$

where $\varphi_0(x)$ is a known, extremely simple function consisting of sections of parallel straight lines, whilst $f_1(x)$ has a Fourier series with coefficients of order not lower than $1/n^2$. We now improve $f_1(x)$. We have:

$$f'(x) = f_1'(x) + m_0.$$

On carrying out the same process on $f'(x)$ as above on $f(x)$, we can write

$$f_1'(x) = f_2(x) + \varphi_1(x),$$

where $\varphi_1(x)$ is a function consisting of pieces of parallel straight lines and $f_2(x)$ has a Fourier expansion with coefficients of order not lower than $1/n^2$. Integration of the last equation gives us an expression for $f_1(x)$, and hence for $f(x)$, as the sum of a Fourier series with coefficients of order not lower than $1/n^3$ and pieces of second degree parabolas. If we were to undertake the further improvement of $f''(x)$, we should get an expression for $f(x)$ as the sum of a Fourier series with coefficients of order not lower than $1/n^4$ and pieces of third degree parabolas, and so on.

The above method is chiefly used when the function is unknown and only its Fourier series given, with coefficients of the form (57). In this case, we have to find the points of discontinuity and the jumps of $f(x)$ and its derivatives from the form of the coefficients, and afterwards apply our method for improving the convergence.

An alternative approach is as follows: we can sum the parts of the Fourier series which derive from the first terms of formulae (57) for the coefficients a_n and b_n. It is with these terms that the poor convergence of the Fourier series is associated. The Fourier series remaining after summation has better convergence than before.

The following formulae must be used for the summation mentioned:

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n} = \begin{cases} \frac{1}{2}(-\pi - x) & (-2\pi < x < 0) \\ \frac{1}{2}\pi - x & (0 < x < 2\pi) \\ 0 & (x = 0 \text{ and } x = \pm 2\pi) \end{cases} \tag{60_1}$$

$$\sum_{n=1}^{\infty} \frac{\cos nx}{n^2} = \begin{cases} \frac{2\pi^2 + 6\pi x + 3x^2}{12} & (-2\pi \leqslant x \leqslant 0) \\ \frac{2\pi^2 - 6\pi x + 3x^2}{12} & (0 \leqslant x \leqslant 2\pi) \end{cases} \tag{60_2}$$

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n^3} = \begin{cases} \frac{2\pi^2 x + 3\pi x^2 + x^3}{12} & (-2\pi \leqslant x \leqslant 0) \\ \frac{2\pi^2 x - 3\pi x^2 + x^3}{12} & (0 \leqslant x \leqslant 2\pi). \end{cases} \tag{60_3}$$

The first formula is obtained by expanding in sines the function $(\pi - x)/2$ in the interval $(0, \pi)$. The second is obtained by integrating the first with respect to x from 0 to x, whilst making use of the equation [144]:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Similarly, the third formula is found by integrating the second. Further integral tion could give us further formulae of the above type. We assume here that the interval has a length equal to π, which can always be arranged for by a simple transformation of the independent variable.

The above idea of improving the convergence of the Fourier series by gradual revision of $f(x)$ and its derivatives, as also the example below, is due to Prof. A. N. Krylov.†

159. Example. We consider the Fourier series:

$$f(x) = -\frac{2}{\pi} \sum_{n=1}^{\infty} \frac{n \cos \frac{n\pi}{2}}{n^2 - 1} \sin nx \qquad (0 \leqslant x \leqslant \pi). \tag{61}$$

We have here:

$$b_n = -\frac{2n \cos \frac{n\pi}{2}}{\pi(n^2 - 1)}.$$

† *O nekotorykh differentsialnykh unavneniyakh matematicheskoi fiziki.*

In order to represent b_n in the form (53), we expand the fraction

$$\frac{n}{n^2-1}$$

in powers of $1/n$, as far as terms of order $1/n^4$:

$$\frac{n}{n^2-1}=\frac{1}{n}+\frac{1}{n^3}+\frac{1}{n^5}\cdot\frac{1}{1-\frac{1}{n^2}}$$

and

$$b_n=-\frac{2\cos\frac{n\pi}{2}}{\pi n}-\frac{2\cos\frac{n\pi}{2}}{\pi n^3}-\frac{2\cos\frac{n\pi}{2}}{\pi n^3(n^2-1)}\,. \tag{62}$$

We thus have to sum the two series:

$$-\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\cos\frac{n\pi}{2}\sin nx}{n}\quad\text{and}\quad-\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\cos\frac{n\pi}{2}\sin nx}{n^3}\,. \tag{63}$$

We let $S_1(x)$ denote the first sum and rewrite it as:

$$S_1(x)=-\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{\sin n\left(x+\frac{\pi}{2}\right)}{n}-\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{\sin n\left(x-\frac{\pi}{2}\right)}{n}\,.$$

Formula (60_1) can be applied to both these sums. We start with the first sum. As x varies from 0 to π, the argument $(x+\pi/2)$ varies from $\pi/2$ to $3\pi/2$, and (60_1) gives:

$$-\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{\sin n\left(x+\frac{\pi}{2}\right)}{n}=-\frac{1}{\pi}\cdot\frac{\pi-\left(x+\frac{\pi}{2}\right)}{2}=\frac{2x-\pi}{4\pi}\quad(0\leqslant x<\pi).$$

As regards the second sum, as x varies from 0 to $\pi/2$, the argument $(x-\pi/2)$ varies from $-\pi/2$ to 0, whilst as x varies from $\pi/2$ to π, $(x-\pi/2)$ varies from 0 to $\pi/2$.

Formula (60_1) gives in this case:

$$-\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{\sin n\left(x-\frac{\pi}{2}\right)}{n}=\begin{cases}\dfrac{2x+\pi}{4\pi} & \left(0\leqslant x<\dfrac{\pi}{2}\right)\\[2ex] \dfrac{2x-3\pi}{4\pi} & \left(\dfrac{\pi}{2}<x\leqslant\pi\right)\\[2ex] 0 & \left(x=\dfrac{\pi}{2}\right).\end{cases}$$

Addition gives us the following final expression for $S_1(x)$:

$$S_1(x) = -\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\cos\frac{n\pi}{2}\sin nx}{n} = \begin{cases} \dfrac{x}{\pi} & \left(0 \leqslant x < \dfrac{\pi}{2}\right) \\ \dfrac{x-\pi}{n} & \left(\dfrac{\pi}{2} < x \leqslant \pi\right) \\ 0 & \left(x = \dfrac{\pi}{2}\right). \end{cases} \tag{64}$$

We could evaluate the second of sums (63) by applying (60_3), but a different approach may be used. Let this sum be denoted as $S_2(x)$. It is easily seen that integrating $S_1(x)$ twice with respect to x gives us $-S_2(x)$ to an accuracy of a first degree polynomial. On integrating (64) twice, we obtain:

$$\frac{x^3}{6\pi}\left(0 \leqslant x < \frac{\pi}{2}\right); \quad \frac{(x-\pi)^3}{6\pi} \quad \left(\frac{\pi}{2} < x \leqslant \pi\right)$$

and consequently:

$$S_2(x) = \begin{cases} -\dfrac{x^3}{6\pi} + C_1' x + C_2' & \left(0 \leqslant x < \dfrac{\pi}{2}\right) \\ -\dfrac{(x-\pi)^2}{6\pi} + C_1'' x + C_2'' & \left(\dfrac{\pi}{2} < x \leqslant \pi\right). \end{cases} \tag{65}$$

We remark in regard to finding the constants C that the Fourier series for $S_2(x)$ has coefficients of order $1/n^3$, whilst the series for $S_2'(x)$ has coefficients of order $1/n^2$; thus both series are uniformly convergent to functions continuous at $x = \pi/2$. It follows that the two expressions (65) must coincide, as must also their derivatives, at $x = \pi/2$:

$$-\frac{\pi^3}{48\pi} + C_1' \cdot \frac{\pi}{2} + C_2' = \frac{\pi^3}{48\pi} + C_1'' \cdot \frac{\pi}{2} + C_2'';$$

$$-\frac{\pi^2}{8\pi} + C_1' = -\frac{\pi^2}{8\pi} + C_1''. \tag{66}$$

Moreover, it follows from the form of the second of sums (63) that $S_2(0) = S_2(\pi) = 0$, which gives by (65):

$$C_2' = 0; \quad C_1''\pi + C_2'' = 0. \tag{67}$$

We can obtain all four constants from equations (66) and (67):

$$C_1' = C_1'' = \frac{\pi}{24}; \quad C_2' = 0; \quad C_2'' = -\frac{\pi^2}{24};$$

and on substituting in (65), we get for $S_2(x)$:

$$S_2(x) = \begin{cases} -\dfrac{x^3}{6\pi} + \dfrac{\pi}{24}x & \left(0 \leqslant x \leqslant \dfrac{\pi}{2}\right) \\ -\dfrac{(x-\pi)^3}{6\pi} + \dfrac{\pi}{24}(x-\pi) & \left(\dfrac{\pi}{2} \leqslant x \leqslant \pi\right). \end{cases}$$

We finally obtain the expression for series (61):

$$f(x) = S_1(x) + S_2(x) - \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\cos \frac{n\pi}{2}}{n^3 (n^2 - 1)} \sin nx, \tag{68}$$

which solves our problem. The function $f(x)$ is given in terms of known functions $S_1(x)$ and $S_2(x)$ consisting of pieces of straight lines and parabolas, and of a Fourier series with coefficients of order

$$\frac{1}{n^3 (n^2 - 1)}, \quad \text{i.e. } \frac{1}{n^5}.$$

We are now free to evaluate the derivatives of the first three orders of $f(x)$, whereas it was impossible to differentiate series (61), which is itself non-uniformly convergent.

§ 16. Fourier integrals and multiple Fourier series

160. Fourier's formula. We conclude our treatment of Fourier series by considering the limiting case, when the interval $(-l, l)$ in which the series is investigated tends to $(-\infty, +\infty)$, i.e. $l \to +\infty$.

Let $f(x)$ satisfy Dirichlet conditions, be continuous in any finite interval, and furthermore be absolutely integrable in $(-\infty, +\infty)$, i.e. there exists:

$$\int_{-\infty}^{+\infty} |f(x)| \, dx = Q.$$

We have by Dirichlet's theorem inside $(-l, l)$:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{l} + b_n \sin \frac{n\pi x}{l} \right).$$

On recalling that

$$a_n = \frac{1}{l} \int_{-l}^{+l} f(t) \cos \frac{n\pi t}{l} \, dt, \quad b_n = \frac{1}{l} \int_{-l}^{+l} f(t) \sin \frac{n\pi t}{l} \, dt,$$

we obtain from this:

$$f(x) = \frac{1}{2l} \int_{-l}^{+l} f(t) \, dt + \frac{1}{l} \sum_{n=1}^{\infty} \int_{-l}^{+l} f(t) \cos \frac{n\pi (t - x)}{l} \, dt.$$

What happens to this formula when $l \to +\infty$? The first term clearly tends to zero, since

$$\left| \frac{1}{2l} \int_{-l}^{+l} f(t)\,\mathrm{d}t \right| \leqslant \frac{1}{2l} \int_{-l}^{+l} |f(t)|\,\mathrm{d}t \leqslant \frac{1}{2l} \int_{-\infty}^{+\infty} |f(t)|\,\mathrm{d}t = \frac{Q}{2l} \to 0 .$$

On introducing a new variable a, which takes equally spaced values in the interval $(0, \infty)$:

$$a_1 = \frac{\pi}{l},\ a_2 = \frac{2\pi}{l},\ \ \ldots,\ a_n = \frac{n\pi}{l},\ \ldots,$$

with each time the increment $\Delta a = \pi/l$, we can write the remaining sum in the form:

$$\frac{1}{\pi} \sum_{(a)} \Delta a \int_{-l}^{+l} f(t) \cos a(t-x)\,\mathrm{d}t .$$

For large l, the integral under the summation sign will only slightly differ from

$$\int_{-\infty}^{+} f(t) \cos a(t-x)\,\mathrm{d}t ,$$

and it can be foreseen that the whole sum will tend to the limit

$$\frac{1}{\pi} \int_0^{\infty} \mathrm{d}a \int_{-\infty}^{+\infty} f(t) \cos a(t-x)\,\mathrm{d}t ,$$

as $l \to +\infty$, so that we have:

$$f(x) = \frac{1}{\pi} \int_0^{+\infty} \mathrm{d}a \int_{-\infty}^{+\infty} f(t) \cos a(t-x)\,\mathrm{d}t . \qquad (1)$$

It is only necessary to replace $f(x)$ at points of discontinuity, if such exist, by

$$\frac{f(x+0) + f(x-0)}{2} .$$

This expression, obtained when $l \to +\infty$ in the Fourier series, is known as Fourier's formula. We have now arrived at the following proposition: *if $f(x)$ satisfies Dirichlet conditions in any finite interval and is absolutely integrable in $(-\infty, +\infty)$, we have the equation, valid for all x:*

$$\frac{1}{\pi} \int_0^{\infty} \mathrm{d}a \int_{-\infty}^{+\infty} f(t) \cos a(t-x)\,\mathrm{d}t = \frac{f(x+0) + f(x-0)}{2} . \qquad (2)$$

The above is generally known as *Fourier's theorem,* whilst the integral on the left-hand side of the equation is the *Fourier integral* of the function $f(x)$. The above discussion is not rigorous, though it can be made so with the help of certain auxiliary arguments. Instead of proceeding in this way, we offer an alternative rigorous proof, based on the results of [152].

Equation (2) will be proved if we can show that

$$\lim_{\lambda\to\infty}\frac{1}{\pi}\int_0^{\lambda} d\alpha \int_{-\infty}^{+\infty} f(t)\cos\alpha(t-x)\,dt = \frac{f(x+0)+f(x-0)}{2}\,.$$

If $J(\lambda, x)$ denotes the integral on the left, we can write:

$$J(\lambda, x) = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(t)\,dt \int_0^{\lambda} \cos\alpha(t-x)\,d\alpha\,, \tag{3}$$

i.e. we can change the order of integrations with respect to t and α.

This follows from the fact that, by the absolute integrability of $f(x)$, the integral

$$\int_{-\infty}^{+\infty} f(t)\cos\alpha(t-x)\,dt \tag{4}$$

is *uniformly* convergent for all values of α. In fact, the integrals

$$\int_N^{N'} f(t)\cos\alpha(t-x)\,dt\,,\quad \int_{-N'}^{-N} f(t)\cos\alpha(t-x)\,dt \qquad (N < N')\ldots \tag{5}$$

do not exceed in absolute value

$$\int_N^{N'} |f(t)|\,dt\,, \tag{6}$$

and consequently, for a given ε an N_0 can be found, independent of α, such that for all N and $N' > N_0$ integrals (5) have absolute values less than ε, this being a property of integral (6), by the absolute integrability of $f(t)$.

But now integral (4) can be integrated with respect to the parameter α under the sign of the integral [84], which gives us:

$$J(\lambda, x) = \frac{1}{\pi}\int_0^{\lambda} d\alpha \int_{-\infty}^{+\infty} f(t)\cos\alpha(t-x)\,dt = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(t)\,dt \int_0^{\lambda} \cos\alpha(t-x)\,d\alpha.$$

The inner integral with respect to α on the right-hand side of (3) can be evaluated directly, which leads us to:

$$J(\lambda, x) = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(t)\,\frac{\sin\lambda(t-x)}{t-x}\,dt\,, \tag{7}$$

so that it remains for us to find:

$$\lim_{\lambda \to \infty} \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \frac{\sin \lambda (t-x)}{t-x} \, dt.$$

On splitting the interval of integration $(-\infty, +\infty)$ into the two intervals $(-\infty, x)$ and $(x, +\infty)$, and replacing $(t-x)$ by $(-z)$ in the first and by z in the second, we can write (7) in the form:

$$J(\lambda, x) = \frac{1}{\pi} \int_0^{\infty} f(x-z) \frac{\sin \lambda z}{z} \, dz + \frac{1}{\pi} \int_0^{\infty} f(x+z) \frac{\sin \lambda z}{z} \, dz.$$

Both these integrals have the form of Dirichlet integrals, except for the infinite limits. In spite of this latter fact, they are easily shown to have the properties of ordinary Dirichlet integrals, i.e. we must have as $\lambda \to \infty$:

$$\left.\begin{aligned} \frac{1}{\pi} \int_0^{\infty} f(x-z) \frac{\sin \lambda z}{z} \, dz &\to \frac{1}{2} f(x-0) \\ \frac{1}{\pi} \int_0^{\infty} f(x+z) \frac{\sin \lambda z}{z} \, dz &\to \frac{1}{\pi} f(x+0), \end{aligned}\right\} \tag{8}$$

whence it follows that in fact:

$$J(\lambda, x) \to \frac{f(x+0) + f(x-0)}{2},$$

which proves Fourier's theorem.

We still have to prove (8). We shall confine ourselves to proving the first of the expressions. Let ε be any given small positive number. With $z > 1$, the factor $(1/z)\sin \lambda z$ has an absolute value less than unity for any real λ, whilst the function $f(x-z)$ is absolutely integrable in $(0, \infty)$ by hypothesis. We can therefore find an $N > 1$ such that, for any λ:

$$\left| \frac{1}{\pi} \int_N^{\infty} f(x-2z) \frac{\sin \lambda z}{z} \, dz \right| < \frac{1}{\pi} \int_N^{\infty} |f(x-2z)| \, dz < \frac{\varepsilon}{2}.$$

If we take the Dirichlet integral with finite limits:

$$\frac{1}{\pi} \int_0^{N} f(x-2z) \frac{\sin \lambda z}{z} \, dz,$$

we can say that it tends to $f(x-0)/2$ as $\lambda \to \infty$, i.e. for all sufficiently large λ:

$$\left| \frac{1}{\pi} \int_0^{N} f(x-2z) \frac{\sin \lambda z}{z} \, dz - \frac{1}{2} f(x-0) \right| < \frac{\varepsilon}{2}.$$

We obviously have:

$$\frac{1}{\pi}\int_0^\infty f(x-2z)\,\frac{\sin\lambda z}{z}\,\mathrm{d}z-\frac{1}{2}f(x-0)=$$

$$=\left[\frac{1}{\pi}\int_0^N f(x-2z)\,\frac{\sin\lambda z}{z}\,\mathrm{d}z-\frac{1}{2}f(x-0)\right]+\frac{1}{\pi}\int_N^\infty f(x-2z)\,\frac{\sin\lambda z}{z}\,\mathrm{d}z\,,$$

whence, by virtue of the last inequalities, we have for all sufficiently large λ:

$$\left|\frac{1}{\pi}\int_0^\infty f(x-2z)\,\frac{\sin\lambda z}{z}\,\mathrm{d}z-\frac{1}{2}f(x-0)\right|<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

In view of the arbitrary smallness of ε, this gives us the first of expressions (8). The proof of the second expression is precisely the same.

A transformation of equation (2) is possible if $f(x)$ is an even or odd function.

We have in fact, on expanding $\cos a(t-x)$ as the cosine of a difference:

$$\frac{f(x+0)+f(x-0)}{2}=$$

$$=\frac{1}{\pi}\int_0^\infty \mathrm{d}a\left[\int_{-\infty}^{+\infty} f(t)\cos at\cos ax\,\mathrm{d}t+\int_{-\infty}^{+\infty} f(t)\sin at\sin ax\,\mathrm{d}t\right], \tag{9}$$

where both integrals with respect to t clearly have a meaning in view of the absolute integrability of $f(t)$ in the interval $(-\infty, +\infty)$.

If $f(t)$ is even, $f(t)\cos at$ is also even, whilst $f(t)\sin at$ is odd, and consequently:

$$\int_{-\infty}^{+\infty} f(t)\cos at\,\mathrm{d}t=2\int_0^\infty f(t)\cos at\,\mathrm{d}t,$$

$$\int_{-\infty}^{+\infty} f(t)\sin at\,\mathrm{d}t=0,$$

so that

$$\frac{f(x+0)+f(x-0)}{2}=\frac{2}{\pi}\int_0^\infty \cos ax\,\mathrm{d}a\int_0^\infty f(t)\cos at\,\mathrm{d}t.$$

On the other hand, if $f(x)$ is odd, we obtain in a similar manner:

$$\frac{f(x+0)+f(x-0)}{2}=\frac{2}{\pi}\int_0^\infty \sin ax\,\mathrm{d}a\int_0^\infty f(t)\sin at\,\mathrm{d}t.$$

If $f(x)$ is defined in $(0, \infty)$ only, it can be contined into the neighbouring interval $(-\infty, 0)$ in an even or odd manner, in which case we get two expressions for the same $f(x)$, assumed for simplicity to be continuous:

$$f(x) = \frac{2}{\pi} \int_0^\infty \cos \alpha x \, d\alpha \int_0^\infty f(t) \cos \alpha t \, dt \quad (x > 0), \tag{10}$$

$$f(x) = \frac{2}{\pi} \int_0^\infty \sin \alpha x \, d\alpha \int_0^\infty f(t) \sin \alpha t \, dt \quad (x > 0). \tag{11}$$

It need only be borne in mind that with the first of these the $f(x)$, contined evenly, gives a continuous function of x, so that the first expression is also valid for $x = 0$; whereas we get a discontinuity with the second expression if $f(0) \neq 0$, the right-hand side being equal to zero, and not $f(0)$, at $x = 0$.

The first integration in (9) is with respect to t, and on introducing the two functions

$$A(\alpha) = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \cos \alpha t \, dt; \quad B(\alpha) = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(t) \sin \alpha x \, dt,$$

we can rewrite (9) in the form:

$$f(x) = \int_0^\infty [A(\alpha) \cos \alpha x + B(\alpha) \sin \alpha x] \, d\alpha,$$

(x) being assumed continuous for simplicity. This latter expression gives us the expansion of $f(x)$ in the infinite interval $(-\infty, +\infty)$ into harmonic oscillations with frequencies α varying continuously from 0 to $+\infty$; the functions $A(\alpha)$ and $B(\alpha)$ give the amplitude distribution laws and the initial phases in relation to the frequency α. For the finite interval $(-l, l)$, we had the frequencies $\alpha_n = n\pi/l$ $(n = 0, 1, \ldots)$, forming an arithmetic progression.

If in (10) we put

$$f_1(\alpha) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(t) \cos \alpha t \, dt, \tag{12$_1$}$$

we can now write (10) as:

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty f_1(\alpha) \cos \alpha x \, d\alpha. \tag{12$_2$}$$

In these two formulae, $f(x)$ and $f_1(a)$ are expressed in terms of each other in precisely the same way.

If we take $f(x)$ as the given, and $f_1(a)$ as the required function in (12_2), this expression now represents a so-called *integral equation* for $f_1(a)$, inasmuch as this latter function appears under the integral sign (Fourier's integral equation). Equation (12_1) gives the solution of this integral equation. Similarly, we can write (11) in the form of the two expressions:

$$f_1(a) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(t) \sin at \, dt \tag{13_1}$$

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty f_1(a) \sin ax \, da, \tag{13_2}$$

Examples. **1.** In expression (10), we put

$$f(x) = \begin{cases} 1 & \text{for } 0 \leqslant x < 1 \\ 0 & \text{,, } \quad x > 1. \end{cases}$$

We now get for the integral on the right-hand side of the equation:

$$\int_0^\infty \cos ax \, da \int_0^\infty f(t) \cos at \, dt = \int_0^\infty \cos ax \, da \int_0^1 \cos at \, dt = \int_0^\infty \frac{\cos ax \sin a}{a} \, da,$$

and consequently

$$\frac{2}{\pi} \int_0^\infty \frac{\cos ax \sin a}{a} \, da = \begin{cases} 1 & \text{for } 0 \leqslant x < 1 \\ \frac{1}{2} & \text{for } x = 1 \\ 0 & \text{for } x > 1. \end{cases}$$

2. On setting in (11)

$$f(x) = e^{-\beta x} \, (\beta > 0),$$

we have on the right-hand side the integral:

$$\frac{2}{\pi} \int_0^\infty \sin ax \, da \int_0^\infty e^{-\beta t} \sin at \, dt = \frac{2}{\pi} \int_0^\infty \frac{a \sin ax}{a^2 + \beta^2} \, da$$

and we thus get

$$\int_0^\infty \frac{a \sin ax}{a^2 + \beta^2} \, da = \begin{cases} \frac{1}{2} \pi e^{-\beta x} & \text{for } x > 0 \\ 0 & \text{for } x = 0. \end{cases}$$

3. Similarly, on setting in (10):

$$f(x) = e^{-\beta x} \ (\beta > 0),$$

we find:

$$\int_0^\infty \frac{\cos \alpha x}{\alpha^2 + \beta^2} \, d\alpha = \frac{\pi}{2\beta} e^{-\beta x}.$$

Fourier's formula is often written in the complex form:

$$\frac{f(x+0) + f(x-0)}{2} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} f(t) \, e^{\alpha(t-x)i} \, dt. \tag{14}$$

This is easily obtained from (2). If we substitute under the integral in (14):

$$e^{\alpha(t-x)i} = \cos \alpha (t-x) + i \sin \alpha (t-x),$$

we get the two integrals:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} f(t) \cos \alpha (t-x) \, dt \text{ and } \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} f(t) \sin \alpha (t-x) \, dt.$$

In the second of these the variable α appears under the sine, so that the integrand is an odd function of α, and we consequently get zero on integrating with respect to α over $(-\infty, +\infty)$. Conversely, we have an even function of α in the first integral, and integration over $(-\infty, +\infty)$ with respect to α can be replaced by integration over $(0, \infty)$ with the factor 2 written in front. Hence it is clear that (14) is equivalent to formula (2).

Assuming continuity of $f(x)$, we rewrite (14) as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\alpha x i} \, d\alpha \int_{-\infty}^{+\infty} f(t) \, e^{\alpha t i} \, dt,$$

whence it is evident that, as in the case of (10) and (11), it can be written in the form of the following two expressions:

$$f_1(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) \, e^{\alpha t i} \, dt, \tag{15$_1$}$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f_1(\alpha) \, e^{-\alpha x i} \, d\alpha. \tag{15$_2$}$$

We notice one point in connection with the complex form of the Fourier integral. We cannot say that

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} f(t)\sin\alpha(t-x)\,dt$$

with infinite limits with respect to the variable α [82] has an ordinary meaning. We can only say that, for any finite positive M,

$$\frac{1}{2\pi}\int_{-M}^{+M} d\alpha \int_{-\infty}^{+\infty} f(t)\sin\alpha(t-x)\,dt = 0,$$

and we should therefore strictly write Fourier's formula in the complex form as

$$f(x) = \frac{1}{2\pi}\lim_{M\to+\infty}\int_{-M}^{+M} e^{-\alpha xi}\,d\alpha \int_{-\infty}^{+\infty} f(t)\,e^{\alpha ti}\,dt.$$

Here, the lower limit tends to $(-\infty)$ and the upper limit to $(+\infty)$, whilst both have the same absolute value. A necessary condition for the existence of the improper integral in the ordinary sense is that a limit exists for any method of the lower limit tending to $(-\infty)$ and the upper to $(+\infty)$.

161. Fourier series in complex form. The method just described for writing the Fourier integral in complex form can likewise be applied to Fourier series.

We recall the formulae of [146]:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\left(a_k\cos\frac{k\pi x}{l} + b_k\sin\frac{k\pi x}{l}\right), \tag{16}$$

$$a_s = \frac{1}{l}\int_{-l}^{+l} f(\xi)\cos\frac{k\pi\xi}{l}\,d\xi;\quad b_k = \frac{1}{l}\int_{-l}^{+l} f(\xi)\sin\frac{k\pi\xi}{l}\,d\xi.$$

We show that these are equivalent to the following:

$$f(x) = \sum_{n=-\infty}^{+\infty} c_n\,e^{i\frac{n\pi x}{l}};\quad c_n = \frac{1}{2l}\int_{-l}^{+l} f(\xi)\,e^{-i\frac{n\pi\xi}{l}}\,d\xi. \tag{17}$$

Here the subscript n takes negative as well as positive integral values. We define separately c_0, c_k and c_{-k}, where k is a positive integer. We have by (17) and (16):

$$c_0 = \frac{1}{2l}\int_{-l}^{+l} f(\xi)\,d\xi = \frac{a_0}{2},$$

$$c_k = \frac{1}{2l}\int_{-l}^{+l} f(\xi)\left(\cos\frac{k\pi\xi}{l} - i\sin\frac{k\pi\xi}{l}\right)d\xi = \frac{a_k - ib_k}{2},$$

$$c_{-k} = \frac{1}{2l}\int_{-l}^{+l} f(\xi)\left(\cos\frac{k\pi\xi}{l} + i\sin\frac{k\pi\xi}{l}\right)d\xi = \frac{a_k + ib_k}{2}.$$

On substituting in series (17) and summing separately over the positive and negative subscripts, we get:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\frac{a_k - ib_k}{2}e^{i\frac{k\pi x}{l}} + \sum_{k=1}^{\infty}\frac{a_k + ib_k}{2}e^{-i\frac{k\pi x}{l}}.$$

Terms in the same k in the two sums written are complex conjugates; on combining the pairs into single terms, we get real quantities:

$$\frac{a_k - ib_k}{2}e^{i\frac{k\pi x}{l}} + \frac{a_k + ib_k}{2}e^{-i\frac{k\pi x}{l}} = a_k\cos\frac{k\pi x}{l} + b_k\sin\frac{k\pi x}{l},$$

and the above expression for $f(x)$ coincides with Fourier series (16) from which the equivalence of (17) and (16) follows.

162. Multiple Fourier series. Fourier series and integrals can also be used for representing functions of two or more independent variables. For instance, let $f(x, y)$ be a periodic function, of period $2l$ with respect to x and $2m$ with respect to y. On considering $f(x, y)$ as a function of x, we have [161]:

$$f(x, y) = \sum_{\sigma=-\infty}^{+\infty} c_\sigma(y)\,e^{i\frac{\sigma\pi x}{l}}, \tag{18}$$

where

$$c_\sigma(y) = \frac{1}{2l}\int_{-l}^{+l} f(\xi, y)\,e^{-i\frac{\sigma\pi\xi}{l}}\,d\xi.$$

The function $c_\sigma(y)$ can in turn be expanded into a series:

$$c_\sigma(y) = \sum_{\tau=-\infty}^{+\infty} c_{\sigma\tau}\,e^{i\frac{\tau\pi y}{m}},$$

where

$$c_{\sigma\tau} = \frac{1}{2m}\int_{-m}^{+m} c_\sigma(\eta)\, e^{-i\frac{\tau\pi\eta}{m}}\, d\eta = \frac{1}{4lm}\int_{-l}^{+l}\int_{-m}^{+m} f(\xi, \eta)\, e^{-i\pi\left(\frac{\sigma\xi}{l}+\frac{\tau\eta}{m}\right)}\, d\xi\, d\eta. \quad (19)$$

Substitution of the expression obtained for $c_\sigma(y)$ in (18) gives us

$$f(x, y) = \sum_{\sigma=-\infty}^{+\infty}\left(\sum_{\tau=-\infty}^{+\infty} c_{\sigma\tau}{}^{i\frac{\tau\pi y}{m}}\right) e^{i\frac{\sigma\pi x}{l}},$$

whence we obtain, on removing the brackets:

$$f(x, y) = \sum_{\sigma, \tau=-\infty}^{\infty} c_{\sigma\tau}\, e^{i\pi\left(\frac{\sigma x}{l}+\frac{\tau y}{m}\right)}, \quad (20)$$

which is the generalization of the Fourier series for the case of two variables.

Similarly, we have for the periodic function $f(x_1, x_2, x_3)$ of three independent variables, of period $2\omega_1$ with respect to x_1, $2\omega_2$ with respect to x_2, and $2\omega_3$ with respect to x_3:

$$f(x_1, x_2, x_3) = \sum_{\sigma_1, \sigma_2, \sigma_3=-\infty}^{+\infty} c_{\sigma_1\sigma_2\sigma_3}\, e^{i\pi\left(\frac{\sigma_1 x_1}{\omega_1}+\frac{\sigma_2 x_2}{\omega_2}+\frac{\sigma_3 x_3}{\omega_3}\right)}, \quad (21)$$

where

$$c_{\sigma_1\sigma_2\sigma_3} = \frac{1}{8\omega_1\omega_2\omega_3}\int_{-\omega_1}^{\omega_1}\int_{-\omega_2}^{\omega_2}\int_{-\omega_3}^{\omega_3} f(\xi_1, \xi_2, \xi_3)\, e^{-i\pi\left(\frac{\sigma_1\xi_1}{\omega_1}+\frac{\sigma_2\xi_2}{\omega_2}+\frac{\sigma_3\xi_3}{\omega_3}\right)}\, d\xi_1\, d\xi_2\, d\xi_3. \quad (22)$$

On separating out the real parts of (20) and (21), we get the real forms of the Fourier expansions.

Series (20) will now take the form, when $l = m = \pi$:

$$f(x, y) = \sum_{n, m=0}^{\infty}(a_{n,m}^{(1)} \cos nx \cos my + a_{n,m}^{(2)} \cos nx \sin my +$$
$$+ a_{n,m}^{(3)} \sin nx \cos my + a_{n,m}^{(4)} \sin nx \sin my).$$

We shall not write down the expressions for the coefficients or investigate the conditions for $f(x, y)$ to have a Fourier expansion. We shall merely note the sufficient condition for expansion to be possible at a given point (x_0, y_0): (1) bounded partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ exist everywhere; (2) the mixed derivative $\partial^2 f/\partial x \partial y$ exists in the

neighbourhood of the point (x_0, y_0), at which it is continuous. Fourier's formula has the form, for functions of two variables:

$$f(x, y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty} \mathrm{d}\alpha_1 \int_{-\infty}^{+\infty} \mathrm{d}\xi \int_{-\infty}^{+\infty} \mathrm{d}\alpha_2 \int_{-\infty}^{+\infty} f(\xi, \eta)\, \mathrm{e}^{i[\alpha_1(\xi - x) + \alpha_2(\eta - y)]}\, \mathrm{d}\eta\,, \quad (23)$$

where the integrations with respect to α_1 and α_2 have to be carried out as indicated at the end of [160]. We have in the real form:

$$f(x, y) = \frac{1}{\pi^2} \int_{0}^{\infty} \mathrm{d}\alpha_1 \int_{-\infty}^{+\infty} \mathrm{d}\xi \int_{0}^{\infty} \mathrm{d}\alpha_2 \int_{-\infty}^{+\infty} f(\xi, \eta) \cos \alpha_1 (\xi - x) \cos \alpha_2 (\eta - y)\, \mathrm{d}\eta. \quad (24)$$

This formula is valid if the function $f(x, y)$, defined throughout the plane, is continuous, has first order partial derivatives, is absolutely integrable with respect to x in $-\infty < x < +\infty$ for any fixed y, and is absolutely integrable with respect to y in $-\infty < y < +\infty$ for any fixed x.

If, for instance, $f(x, y)$ is an even function of x and y, we can write instead of (24):

$$f(x, y) = \frac{4}{\pi^2} \int_{0}^{\infty} \cos \alpha_1 x\, \mathrm{d}\alpha_1 \int_{0}^{\infty} \cos \alpha_1 \xi\, \mathrm{d}\xi \int_{0}^{\infty} \cos \alpha_2 \eta\, \mathrm{d}\alpha_2 \int_{0}^{\infty} f(\xi, \eta) \cos \alpha_2 \eta\, \mathrm{d}\eta. \quad (25)$$

Similarly, Fourier's formula can be written for a function $f(x_1, x_2, \ldots, x_n)$ of any number of independent variables.

CHAPTER VII

THE PARTIAL DIFFERENTIAL EQUATIONS OF MATHEMATICAL PHYSICS

§ 17. The wave equation

163. The equation of vibration of a string. The problem of integrating partial differential equations belongs to one of the most difficult and extensive sections of analysis and we confine ourselves to the consideration of certain particular problems in this field. The present article is devoted to problems connected with the so-called wave equation:

$$\frac{\partial^2 u}{\partial t^2} = a^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) \quad \text{or} \quad \frac{\partial^2 u}{\partial t^2} = a^2 \Delta u .$$

where

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \operatorname{div}\operatorname{grad} u .$$

We arrived at this equation in [116] and [118] when considering acoustic and electromagnetic oscillations. We suppose that u does not depend on y and z, i.e. that u has the same value at all points of a plane perpendicular to the x axis. In this case, the wave equation takes the form

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2},$$

the wave concerned being generally referred to as a plane wave. We now show that the same equation is obtained by considering the small transverse vibrations of a taut homogeneous string.

We understand by string a thin thread which can bend freely. We take it as acted on by a tension T_0 and directed along the x axis when in the equilibrium position with no external forces (Fig. 127). If we move it from the equilibrium position and moreover subject it to the action of a given force, it starts to vibrate, so that a point of the string with the equilibrium position N of abscissa x occupies the position M

at the instant t. We confine ourselves to considering only *transverse vibrations*, on the assumption that all the motion takes place in a single plane and that points of the string move perpendicularly to the x axis. We let u denote the displacement $\overline{NM}$ of a point of the string. This displacement is in fact the required function of the two independent variables x and t.

We distinguish an element MM' of the string, which has the equilibrium position NN'. On the assumption of small deformations, we neglect the square of the derivative $\partial u/\partial x$ by comparison with unity. Let α be the acute angle that the direction of the tangent to the string forms with the x axis. We have:

$$\tan\alpha = \frac{\partial u}{\partial x} \quad \text{and} \quad \sin\alpha = \frac{\tan\alpha}{\sqrt{1+\tan^2\alpha}} = \frac{\dfrac{\partial u}{\partial x}}{\sqrt{1+\left(\dfrac{\partial u}{\partial x}\right)^2}} \sim \frac{\partial u}{\partial x}.$$

Let F denote the force acting on the string perpendicularly to the x axis, and reckoned per unit length. The following forces act on our element MM': the tension at the point M', directed along the tangent at M' and forming an acute angle with the x axis, the tension at the point M, directed along the tangent at M and forming an obtuse angle with the x axis, and the force $F\mathrm{d}x$, directed along the u axis. In view of our assumption of small deformations, we can take both the above tensions as equal in magnitude to T_0. We suppose that initially we have equilibrium of the string under the action of the force F. On projecting on to the u axis, we have the following equilibrium condition:

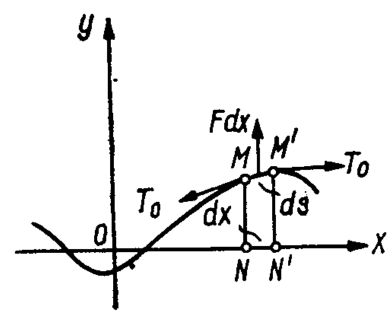

FIG. 127

$$T_0\sin\alpha' - T_0\sin\alpha + F\mathrm{d}x = 0\,, \tag{1}$$

where α' is the value of the angle α at the point M', i.e.

$$\sin\alpha' = \left(\frac{\partial u}{\partial x}\right)_{M'}; \quad \sin\alpha = \left(\frac{\partial u}{\partial x}\right)_M,$$

and consequently:

$$T_0\left[\left(\frac{\partial u}{\partial x}\right)_{M'} - \left(\frac{\partial u}{\partial x}\right)_M\right] + F\mathrm{d}x = 0. \tag{2}$$

The difference in square brackets expresses the increment of the function $\partial u/\partial x$ as a result of x changing by $\mathrm{d}x$. On replacing this increment by the differential, we get [I, 50]:

$$\left(\frac{\partial u}{\partial x}\right)_{M'} - \left(\frac{\partial u}{\partial x}\right)_{M} = \frac{\partial^2 u}{\partial x^2}\,\mathrm{d}x.$$

On substituting in (2) and cancelling $\mathrm{d}x$, we arrive at the *equilibrium equation of the string:*

$$T_0 \frac{\partial^2 u}{\partial x^2} + F = 0. \tag{3}$$

In accordance with d'Alembert's principle, the *equation of motion* is simply obtained by adding to the external force the *inertia force,* which we find as follows: the velocity of the point M is evidently $\partial u/\partial t$, and its acceleration $\partial^2 u/\partial t^2$, so that the inertia force of the element MM', given by the product of the acceleration and the mass taken with the reversed sign, is

$$-\frac{\partial^2 u}{\partial t^2}\,\varrho\,\mathrm{d}x\,,$$

where ϱ is the longitudinal density of the string, i.e. the mass per unit length; the inertia force, reckoned per unit length, is now

$$-\varrho \frac{\partial^2 x}{\partial t^2}\,,$$

ϱ being assumed constant.

It follows that the equation of motion is obtained by replacing F in (3) by $F - \varrho\partial^2 u/\partial t^2$, which gives:

$$\varrho \frac{\partial^2 u}{\partial t^2} = T_0 \frac{\partial^2 u}{\partial x^2} + F.$$

On dividing by ϱ and writing

$$\frac{T_0}{\varrho} = a^2, \quad \frac{F}{\varrho} = f, \tag{4}$$

we obtain the *equation of forced transverse vibration of a string:*

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} + f. \tag{5}$$

If external force is absent, we have $f = 0$, and we obtain for the *equation of free vibration of a string:*

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}. \tag{6}$$

Another method of deducing (5), on the basis of Hamilton's principle, will be found in Volume IV.

We assumed above that the external force was distributed continuously along the length of the string; but occasionally, we are concerned with a force P concentrated at a single point C. This case can be considered, either as a limiting case of the above, the force being assumed to act on an infinitesimal element of length ε about the point C, whereas the product of its magnitude with ε tends to a finite non-zero limit as $\varepsilon \to 0$, or else directly, by applying equation (2) to the element MM' about the point C and replacing the $F dx$ appearing there by P. It may be noticed that we do not add to $F dx$ the inertia force $(-\,(\partial^2 u/\partial t^2)\,\varrho\,dx)$, since we assume that the latter tends to zero as $dx \to 0$.

On assuming that the ends of the element approach the point C, and denoting the respective limiting values to which $\partial u/\partial x$ tends as we approach C from the right and the left by

$$\left(\frac{\partial u}{\partial x}\right)_+,\ \left(\frac{\partial u}{\partial x}\right)_-,$$

equation (2) gives us in the limit:

$$T_0\left[\left(\frac{\partial u}{\partial x}\right)_+ - \left(\frac{\partial u}{\partial x}\right)_-\right] = -P. \tag{7}$$

We see from this that the string has an angular point at the point C of application of the concentrated force, i.e. a point with different tangential directions to the right and the left.

As is generally the case in dynamics, the equation of motion (5) is not sufficient in itself for a full definition of the motion of the string; we also have to give the state at the initial instant $t = 0$, that is, the positions of its points u and their velocities $\partial u/\partial t$ at $t = 0$ must be known functions of x:

$$u\Big|_{t=0} = \varphi(x); \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \varphi_1(x). \tag{8}$$

These conditions, which have to be satisfied by the required function u at $t = 0$, are known as *initial conditions.*

An infinite string may be considered in theory; equation (5) and conditions (8) are sufficient for finding the solution here, where $\varphi(x)$ and $\varphi_1(x)$ have to be specified throughout the infinite interval $(-\infty, +\infty)$. This case can correspond to the consideration of plane waves in unbounded space. As we shall see later, the results obtained for an

infinite string give us a picture of the distribution of disturbances in a bounded string also, up to the instant when disturbances reflected from the ends also make their appearance at the point concerned.

If, however, *the string is bounded on one or both sides* at the points $x = 0$ and $x = l$, it is necessary to indicate what happens at the ends. For instance, let the end $x = 0$ be fixed. We must have in this case

$$u\,|_{x=0} = 0. \tag{9}$$

If the end $x = l$ is also fixed, we likewise have:

$$u\,|_{x=l} = 0, \tag{9_1}$$

and these conditions must be fulfilled for any t.

Instead of the ends being fixed, they may move in a given manner; the ordinates of the ends must now be given functions of time, i.e. we must have:

$$u\,|_{x=0} = \chi_1(t)\,; \quad u\,|_{x=l} = \chi_2(t). \tag{10}$$

Whatever the case, if the string is bounded on one or both sides, a condition must be a specified for each end, this being known as a *boundary condition.*

We see from the above that *the initial and boundary conditions have as great an importance as the actual equation of motion for the solution of a concrete physical problem,* and moreover, that what we are interested in is not so much finding arbitrary solutions or even the general solution as finding those solutions which satisfy the given initial and boundary conditions.

164. D'Alembert's solution. In the case of the free vibration of an infinite string, the required function $u(x, t)$ must satisfy equation (6):

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}$$

with initial conditions (8):

$$u\,|_{t=0} = \varphi(x)\,; \quad \left.\frac{\partial u}{\partial t}\right|_{t=0} = \varphi_1(x)\,,$$

where $\varphi(x)$ and $\varphi_1(x)$ must be specified in the interval $(-\infty, +\infty)$ since the string is infinite.

The general solution of (6) can in fact be found, and in such a form that conditions (8) can easily be satisfied.

For this, we transform (6) to the new independent variables:

$$\xi = x - at\,, \quad \eta = x + at$$

or

$$x = \frac{1}{2}(\eta + \xi); \quad t = \frac{1}{2a}(\eta - \xi).$$

On taking u as depending on x and t indirectly via ξ and η and using the rule for differentiating functions of a function, we can express the derivatives with respect to the first variables in terms of the derivatives with respect to the new variables:

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial \xi} + \frac{\partial u}{\partial \eta}; \quad \frac{\partial u}{\partial t} = a\left(\frac{\partial u}{\partial \eta} - \frac{\partial u}{\partial \xi}\right).$$

On applying these formulae a second time, we get:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial \xi}\left(\frac{\partial u}{\partial \xi} + \frac{\partial u}{\partial \eta}\right) + \frac{\partial}{\partial \eta}\left(\frac{\partial u}{\partial \xi} + \frac{\partial u}{\partial \eta}\right) = \frac{\partial^2 u}{\partial \xi^2} + 2\frac{\partial^2 u}{\partial \xi\,\partial \eta} + \frac{\partial^2 u}{\partial \eta^2},$$

$$\frac{\partial^2 u}{\partial t^2} = a^2\frac{\partial}{\partial \eta}\left(\frac{\partial u}{\partial \eta} - \frac{\partial u}{\partial \xi}\right) - a^2\frac{\partial}{\partial \xi}\left(\frac{\partial u}{\partial \eta} - \frac{\partial u}{\partial \xi}\right) = a^2\left(\frac{\partial^2 u}{\partial \xi^2} - 2\frac{\partial^2 u}{\partial \xi\,\partial \eta} + \frac{\partial^2 u}{\partial \eta^2}\right),$$

whence

$$\frac{\partial^2 u}{\partial t^2} - a^2\frac{\partial^2 u}{\partial x^2} = -4a^2\frac{\partial^2 u}{\partial \xi\,\partial \eta},$$

and equation (6) is seen to be equivalent to the following:

$$\frac{\partial^2 u}{\partial \xi\,\partial \eta} = 0. \tag{11}$$

We conclude, on re-writing (11) in the form

$$\frac{\partial}{\partial \eta}\left(\frac{\partial u}{\partial \xi}\right) = 0,$$

that $\partial u/\partial \xi$ is independent of η, i.e. is a function of ξ only. If we write

$$\frac{\partial u}{\partial \xi} = \theta(\xi),$$

integration gives us:

$$u = \int \theta(\xi)\,d\xi + \theta_2(\eta),$$

where $\theta_2(\eta)$ is an arbitrary function of η (the "constant" of integration with respect to ξ can depend on η). The first term can be reckoned here as an arbitrary function of ξ, since $\theta(\xi)$ is an arbitrary function of ξ; on writing the first term as $\theta_1(\xi)$, we have:

$$u = \theta_1(\xi) + \theta_2(\eta),$$

or, on returning to the old variables x and t:

$$u(x, t) = \theta_1(x - at) + \theta_2(x + at), \tag{12}$$

where θ_1 and θ_2 are arbitrary functions of their arguments. This is the most general solution of (6) and is known as d'Alembert's solution; it contains two arbitrary functions θ_1 and θ_2. We define these by bringing in the initial conditions (8), which, in view of the equation

$$\frac{\partial u}{\partial t} = a\left[-\theta_1'(x - at) + \theta_2'(x + at)\right]$$

and equation (12), give:

$$\theta_1(x) + \theta_2(x) = \varphi(x)\,; \quad -\theta_1'(x) + \theta_2'(x) = \frac{\varphi_1(x)}{a} \tag{13}$$

or, on integrating and reversing the signs:

$$\theta_1(x) - \theta_2(x) = -\frac{1}{a}\int_0^x \varphi_2(z)\,\mathrm{d}z + C.$$

We determine the arbitrary constant C by putting $x = 0$:

$$C = \theta_1(0) - \theta_2(0).$$

We can take $C = 0$ without loss of generality, i.e.

$$\theta_1(0) - \theta_2(0) = 0\,, \tag{14}$$

since, if we happened to have $C \neq 0$, we could replace $\theta_1(x)$ and $\theta_2(x)$ by the functions:

$$\theta_1(x) - \frac{C}{2}\,, \quad \theta_2(x) + \frac{C}{2}\,,$$

so that (14) is satisfied whilst equations (13) are left unchanged.

Thus we have:

$$\theta_1(x) + \theta_2(x) = \varphi(x)\,; \quad \theta_1(x) - \theta_2(x) = -\frac{1}{a}\int_0^x \varphi_1(z)\,\mathrm{d}z. \tag{15}$$

From these, we can easily find $\theta_1(x)$ and $\theta_2(x)$:

$$\theta_1(x) = \frac{1}{2}\varphi(x) - \frac{1}{2a}\int_0^x \varphi_1(z)\,\mathrm{d}z\,; \quad \theta_2(x) = \frac{1}{2}\varphi(x) + \frac{1}{2a}\int_0^x \varphi_1(z)\,\mathrm{d}z. \tag{16}$$

On subsituting the expressions obtained in (12), we find:

$$u(x, t) = \frac{1}{2}\varphi(x - at) - \frac{1}{2a}\int_0^{x-at} \varphi_1(z)\,\mathrm{d}z + \\ + \frac{1}{2}\varphi(x + at) + \frac{1}{2a}\int_0^{x+at} \varphi_1(z)\,\mathrm{d}z,$$

or finally:

$$u(x,t) = \frac{\varphi(x-at)+\varphi(x+at)}{2} + \frac{1}{2a}\int\limits_{x-at}^{x+at} \varphi_1(z)\,dz \tag{17}$$

Expression (17) clearly gives a twice continuously differentiable solution (the so-called classical solution) of the problem, if $\varphi(x)$ has continuous derivatives $\varphi'(x)$ and $\varphi''(x)$, whilst $\varphi_1(x)$ has a continuous derivative $\varphi_1'(x)$ for $-\infty < x < +\infty$. It quite often happens, however, that we are concerned with problems in which the initial disturbance is specified by functions which do not satisfy these conditions. For instance, if the string has a polygonal form at the initial instant, $\varphi(x)$ lacks a definite derivative at the vertices of the polygonal shape. Nevertheless, it is reasonable to suppose that (17) gives the solution of the problem in spite of $u(x, t)$ not having continuous derivatives up to the second order everywhere. The problem is said to have a generalized solution in this case. The theory of generalized solutions is dealt with in Volume IV.

165. Particular cases. Formula (17) gives the full solution of the problem considered. It becomes more clearly comprehensible on distinguishing various particular cases.

1. *The initial impulse is zero*, i.e. the initial velocity of points of the string is zero. With this condition, $\varphi_1(x) = 0$ and (17) gives:

$$u(x,t) = \frac{\varphi(x-at)+\varphi(x+at)}{2}, \tag{18}$$

together with, at the initial instant:

$$u\big|_{t=0} = u(x,0) = \varphi(x).$$

What are the physical interpretations of solution (18)? The numerator in (18) consists of two terms, and we shall dwell on the first, $\varphi(x - at)$.

We suppose that an observer, departing from the point $x = c$ of the string at the initial instant $t = 0$, moves in the positive direction of OX with a velocity a, i.e. his abscissa changes in accordance with the formula: $x = c + at$ or $x - at = c$. For such an observer, the displacement of the string, defined by $u = \varphi(x - at)$, will always remain constant, equal to $\varphi(c)$. The actual phenomenon defined by the function $u = \varphi(x - at)$ is known as *the direct wave propagation*. Returning to d'Alembert's formula (12), we can say that the term $\theta_1(x - at)$ gives

the direct wave which is propagated along the positive direction of OX with velocity a. Similarly, the second term $\theta_2(x + at)$ defines a vibration of the string, the disturbance from which is propagated with velocity a in the negative direction of OX, in the sense that a point with abscissa $c - at$ at the instant t will have the same deviation u as the point $x = c$ had at $t = 0$. We call the corresponding phenomenon the *reverse wave propagation.*

The quantity a is the velocity of propagation of the disturbance or (transverse) vibration. We can see from (4) that

$$a = \sqrt{\frac{T_0}{\varrho}}, \tag{19}$$

i.e. *the velocity of propagation of transverse vibrations is inversely proportional to the square root of the density of the string and directly proportional to the square root of the tension.*

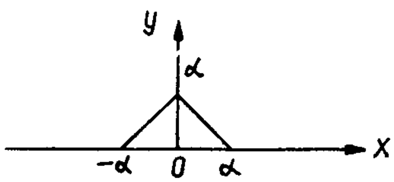

FIG. 128

Solution (18) obtained above represents the arithmetic mean of the direct wave $\varphi(x-at)$ and the reverse wave $\varphi(x+at)$ and can be obtained as follows: we draw two identical samples of the graph $u = \varphi(x)$ of the string at $t = 0$ and imagine that they are superimposed on each other then displaced in either direction with velocity a. The graph of the string at the instant t is obtained as the arithmetic mean of the displaced graphs, i.e. it bisects the ordinates of these latter graphs.

For instance, let the initial shape of the string be as shown in Fig. 128:

$$\varphi(x) = \begin{cases} 0 \text{ outside the interval } (-\alpha, \alpha) \\ x + \alpha \quad \text{for} \quad -\alpha < x < 0 \\ -x + \alpha \quad \text{for} \quad 0 < x < \alpha \end{cases}$$

Figure 129 illustrates the graphs of the string at the instants

$$t = \frac{\alpha}{4a}, \frac{2\alpha}{4a}, \frac{3\alpha}{4a}, \frac{\alpha}{a}, \frac{5\alpha}{4a}, \frac{2\alpha}{a},$$

We take two axes at right angles on the plane: one for the variable x and the other for t. The x axis only is drawn in Fig. 130. Every point of the plane is defined by the two coordinates (x, t), i.e. every point characterizes a definite point of the string x at a definite instant t. We can now readily find by graphical means the points of the string whose initial disturbances arrive at the point x_0 at the instant t_0.

By the above, these will be the points with abscissae $x_0 + at_0$, since a is the velocity of propagation of the vibration. We find these points on the x axis simply by drawing through the point (x_0, t_0) the two straight lines:

$$\left.\begin{aligned} x - at &= x_0 - at_0 \,, \\ x + at &= x_0 + at_0 \,, \end{aligned}\right\} \tag{20}$$

the intersections of these with OX being the required points. Straight lines (20) are known as *characteristics of the point* (x_0, t_0). Along the

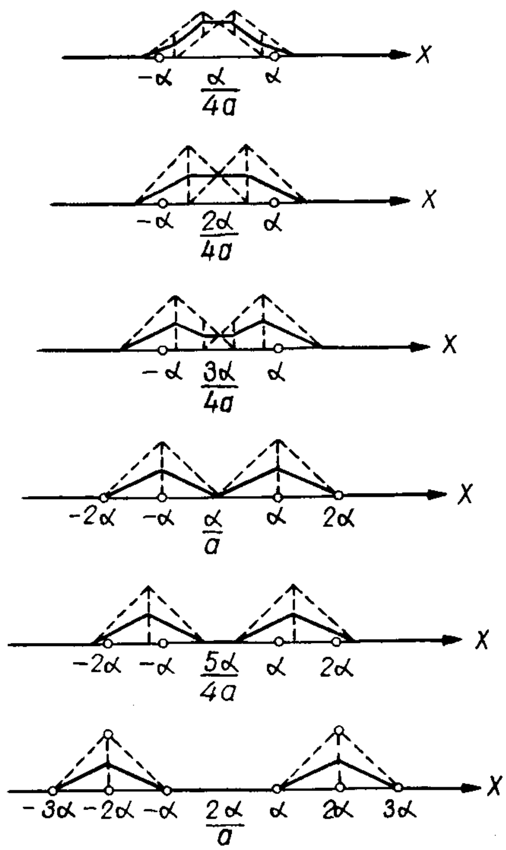

FIG. 129

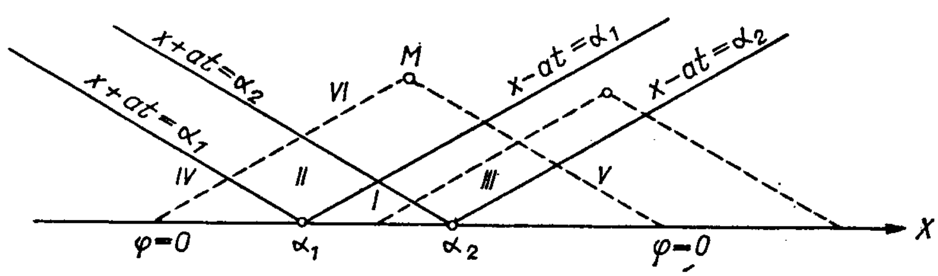

FIG. 130

first of these lines, $\varphi(x - at)$ retains a constant value, i.e. the line gives the values (x, t) for which the direct wave produces the same deviation as at (x_0, t_0). The second of lines (20) plays the same role as regards the reverse wave. We can say briefly that *the disturbances are propagated along the characteristics.*

We can reveal the following facts on applying the construction indicated.

Let an initial disturbance have been present only in an interval (a_1, a_2) of the string (Fig. 130), i.e. $\varphi(x) = 0$ outside this interval. We confine ourselves to the upper part of the (x, t) plane, i.e. $t > 0$, which alone has a physical meaning, and draw the characteristics of the points a_1 and a_2 on OX, these being represented by full lines. These characteristics divide the total half-plane into *six* domains. Domain (I) corresponds to those points at which both direct and reverse waves arrive at the given instant. Domain (II) corresponds to points at which only a reverse wave arrives at the given instant; whereas only direct waves arrive in domain (III). Points of domains (IV) and (V) are those at which no disturbance has arrived up to the given instant. Finally, the disturbance has had time to come and go through points of domain (VI), and these find themselves in a state of rest at the given instant. This is clear from the fact that if characteristics are drawn through any point M of this last domain, they will intersect OX at some point $x = c$ outside the interval of initial disturbance, and consequently the values of $\varphi(x \pm at) = \varphi(c)$ will be zero. Furthermore, if a line is drawn through M perpendicular to OX, the lower part of this line, which corresponds to the earlier instants with fixed x, will intersect at least one of the domains (I), (II) , (III), whilst the upper part of the line, corresponding to later instants, will be situated entirely in domain (VI). This remarkable property, of returning to the original state after passage of the wave, is not possessed by the string with every initial disturbance, as will be seen below.

2. *The initial displacement is zero,* and only an initial impulse is present.

Here we get the solution:

$$u(x, t) = \frac{1}{2a} \int_{x-at}^{x+at} \varphi_1(z)\,dz. \tag{21}$$

If we write $\Phi_1(x)$ for the indefinite integral of $\varphi_1(x)/2a$, we have:

$$u(x, t) = \Phi_1(x + at) - \Phi_1(x - at), \tag{22}$$

i.e. it is again a matter of the propagation of a direct and reverse wave. If the initial disturbance is confined to the interval (a_1, a_2), we get the same construction as in case 1, with the important difference, however, that the displacement in domain (VI) differs from zero and is given by the integral:

$$\frac{1}{2a}\int_{a_1}^{a_2}\varphi_1(z)\,dz. \tag{23}$$

In fact, it follows from the construction of domain (VI) that we have for points of it $x + at > a_2$ and $x - at < a_1$, i.e. the integration in (21) has to be carried out over an interval that contains (a_1, a_2). But $\varphi_1(z)$ is zero outside (a_1, a_2) by hypothesis, so that only the integral over (a_1, a_2) remains, and we obtain expression (23) for $u(x, t)$, which represents a certain constant.

The action of the initial impulse thus amounts to points of the string undergoing displacements in the course of time that are expressed by integral (23), after which the points remain without movement in the new position.

Equation (21) can also be interpreted as follows. Let the point x lie to the right of the interval (a_1, a_2), i.e. $x > a_2$. The interval of integration degenerates to the point x at $t = 0$, then with increasing t, it extends to both sides with velocity a. It will have no points in common with (a_1, a_2) for $t < (x - a_2)/a$, $\varphi_1(z)$ will be zero, and (21) gives $u(x, t) = 0$, i.e. rest at the point x. Starting at the instant $t = (x - a_2)/a$, the interval $(x - at, x + at)$ will overlap with (a_1, a_2), in which $\varphi_1(z)$ differs from zero, and the point x starts to vibrate (the instant when the front of the wave passes through x). Finally, with $t > (x - a_1)/a$ the interval $(x - at, x + at)$ will contain the entire interval (a_1, a_2), integration over $(x - at, x + at)$ reduces to integration over (a_1, a_2) since $\varphi_1(z)$ is zero outside the latter by hypothesis, and we thus have now a constant value for $u(x, t)$, given by (23). The instant $t = (x - a_1)/a$ is when the rear front of the wave passes across the point x.

We notice some details regarding the general case. It can happen in the general case that the direct or reverse wave is entirely absent. Suppose, for instance, that the functions $\varphi(x)$ and $\varphi_1(x)$ appearing in the initial conditions satisfy the relationship:

$$\frac{1}{2}\varphi(x) + \frac{1}{2a}\int_{0}^{x}\varphi_1(z)\,dz = 0. \tag{24}$$

With this, by the second of expressions (16), the function $\theta_2(x)$ is identically zero, and the reverse wave is absent in the general solution (12). If we replace the zero on the right of (24) by a constant, $\theta_2(x)$ becomes constant and its value can be absorbed into $\theta_1(x-at)$ in (12), i.e. reverse wave is again absent. Let us return to the example that we considered in case 1. Figure 128 gives the graph of the initial deviation (the initial velocity is zero everywhere). The last of Figs. 129 gives the graph of the string at the instant $t = t_0$, consisting of two separate pieces. The right-hand piece corresponding to the interval $(a, 3a)$ will move to the right, and the left-hand piece to the left, with velocity a. But we can describe the later phenomena with $t > t_0$ by taking $t = t_0$ as the initial instant, calculating the deviation u and velocity $\partial u/\partial t$ for this instant, and applying the general formula (17), in which it is only necessary to replace t by $(t - t_0)$ on the right-hand side, t_0 having been taken as the initial instant. In this case, the initial conditions only differ from zero in the intervals $(-3a, -a)$ and $(a, 3a)$. In the general case, the disturbances in each of these intervals might give both a direct and a reverse wave. Here, however, as we saw above, the disturbances say in the interval $(a, 3a)$ give only a direct wave. This happens because in this interval, apart from the initial deviations illustrated in the last of Figs. 129, velocities are excited as a result of the vibration at $t = t_0$ such that the reverse wave is absent. Similarly, the disturbances over the piece $(-3a, -a)$ do not give a direct wave. This phenomenon corresponds to one of the formulations of Huygens principle.

166. Finite string. Let the string be finite and fixed at the ends and let the abscissae of the ends be $x = 0$ and $x = l$.

In addition to the initial conditions (8):

$$u\Big|_{t=0} = \varphi(x)\,; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \varphi_1(x)\,,$$

where $\varphi(x)$ and $\varphi_1(x)$ are specified for $0 < x < l$, we also have to satisfy the boundary conditions:

$$u\,|_{x=0} = 0\,; \quad u\,|_{x=l} = 0\,. \tag{25}$$

d'Alembert's solution (12):

$$u(x,t) = \theta_1(x-at) + \theta_2(x+at)\,, \tag{12}$$

is of course suitable for this case, but the definitions of the functions θ_1 and θ_2 in accordance with (16):

$$\left.\begin{aligned} \theta_1(x) &= \frac{1}{2}\varphi(x) - \frac{1}{2a}\int_0^x \varphi_1(z)\,\mathrm{d}z\,; \\ \theta_2(x) &= \frac{1}{2}\varphi(x) + \frac{1}{2a}\int_0^x \varphi_1(z)\,\mathrm{d}z\,; \end{aligned}\right\} \tag{26}$$

meet with the difficulty here that $\varphi(x)$ and $\varphi_1(x)$, and consequently also $\theta_1(x)$ and $\theta_2(x)$, are only defined in the interval $(0, l)$ in line with the physical meaning of the problem, whereas the arguments $(x \pm at)$ in (12) can lie outside the interval.

Consequently, for the characteristics method to be applicable we have to continue the functions $\theta_1(x)$ and $\theta_2(x)$, or what is exactly equivalent, $\varphi(x)$ and $\varphi_1(x)$, outside the interval $(0, l)$. From the physical view-point, this continuation amounts to defining the initial disturbance of an *infinite* string such that the motion of its part $(0, l)$ is the same as if it were fixed at the ends and the rest of the string removed.

On substituting $x = 0$ and $x = l$ in the right-hand side of (12), the boundary conditions become, on equating the results to zero:

$$\left.\begin{aligned} \theta_1(-at) + \theta_2(at) &= 0\,; \\ \theta_1(l - at) + \theta_2(l + at) &= 0\,, \end{aligned}\right\} \tag{27}$$

or, if we write the variable argument at simply as x:

$$\left.\begin{aligned} \theta_1(-x) &= -\theta_2(x)\,; \\ \theta_2(l + x) &= -\theta_1(l - x). \end{aligned}\right\} \tag{28}$$

When x varies in $(0, l)$, the argument $(l - x)$ also varies in this interval, and the right-hand sides of equations (28) are known to us. But the arguments $(-x)$ and $(l + x)$ now vary respectively in the intervals $(-l, 0)$ and $(l, 2l)$, and the second of equations (28) gives us values of $\theta_2(x)$ in $(l, 2l)$, whilst the first gives $\theta_1(x)$ in $(-l, 0)$. Furthermore, as x varies in $(l, 2l)$, the argument $(l - x)$ varies in $(-l, 0)$, and the right-hand sides of equations (28) are known to us on the basis of the above working. Arguments $(-x)$ and $(l + x)$ now vary in the intervals $(-2l, -l)$ and $(2l, 3l)$, so that (28) gives us $\theta_2(x)$ in $(2l, 3l)$ and $\theta_1(x)$ in $(-2l, -l)$. We can see by continuing in the same way that (28) gives us the determinate values for $\theta_1(x)$ with $x \leqslant 0$ and $\theta_2(x)$ with $x \geqslant l$, which we require for applying (12) with $t > 0$. Similarly, if x

varies in the interval $(-l, 0)$, the left-hand sides of equations (28) are known, and we obtain $\theta_2(x)$ in $(-l, 0)$ and $\theta_1(x)$ in $(l, 2l)$. With x varying next in $(-2l, -l)$, we get $\theta_2(x)$ in $(-2l, -l)$ and $\theta_1(x)$ in $(2l, 3l)$ and so on, i.e. equations (28) give us $\theta_1(x)$ and $\theta_2(x)$ defined for all real x.

If we replace x by $(l + x)$ in the second of equations (28) and make use of the first equation, we get:

$$\theta_2(x + 2l) = -\theta_1(-x) = \theta_2(x),$$

i.e. the function $\theta_2(x)$ is shown to have the period $2l$. It now follows from the first of (28) that $\theta_1(x)$ also has the period $2l$. This implies that we in fact know $\theta_1(x)$ and $\theta_2(x)$ for all real x provided simply that we carry out the first of the operations described above for continuing these functions, i.e. it is sufficient to let x vary in $(0, l)$ only. Equations (28) give us $\theta_1(x)$ in $(-l, 0)$ and $\theta_2(x)$ in $(l, 2l)$, i.e. $\theta_1(x)$ is known in $(-l, +l)$ whilst $\theta_2(x)$ is known in $(0, 2l)$. The remaining values of the functions follow from their periodicity.

Having thus defined $\theta_1(x)$ and $\theta_2(x)$, the functions $\varphi(x)$, $\varphi_1(x)$ are readily continued, since we have by equations (26):

$$\varphi(x) = \theta_1(x) + \theta_2(x); \quad \frac{1}{a}\int_0^x \varphi_1(z)\,dz = \theta_2(x) - \theta_1(x),$$

i.e.

$$\varphi_1(x) = a[\theta_2'(x) - \theta_1'(x)].$$

We obtain on replacing x by $(-x)$ in the first of equations (28) and on differentiating also:

$$\theta_1(x) = -\theta_2(-x); \quad \theta_1'(-x) = \theta_2'(x); \quad \theta_1'(x) = \theta_2'(-x).$$

We use these relationships and the first of equations (28) to enable us to write:

$$\varphi(-x) = \theta_1(-x) + \theta_2(-x) = -\theta_2(x) - \theta_1(x) = -\varphi(x)$$
$$\varphi_1(-x) = a\left[\theta_2'(-x) - \theta_1'(-x)\right] = a\left[\theta_1'(x) - \theta_2'(x)\right] = -\varphi_1(x),$$

i.e. we obtain an extremely simple rule for continuing $\varphi(x)$ and $\varphi_1(x)$: they are continued from the interval $(0, l)$ to $(-l, 0)$ oddly, and thereafter periodically with period $2l$. If, with this, we obtain functions $\varphi(x)$ and $\varphi_1(x)$ along the entire x axis such that $\varphi(x)$ has continuous derivatives $\varphi'(x)$ and $\varphi''(x)$, whilst $\varphi_1(x)$ has a continuous derivative $\varphi_1'(x)$, we shall have by (17) a twice continuously differentiable solution of our problem.

We return once more to the xt plane. Since the string is finite, we need only consider the strip of the upper half-plane $t > 0$ included between the lines $x = 0$, $x = l$ (Fig. 131). The physical significance of the solution (12), in which the functions $\theta_1(x)$ and $\theta_2(x)$ are defined for all x, as shown above, may be explained as follows. Having drawn characteristics through the points O and L to their intersections with the opposite boundaries of the strip, we draw new characteristics through these points of intersection to their further intersections with opposite boundaries of the strip, and so on.

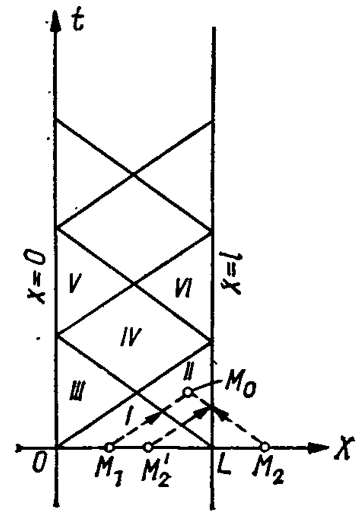

FIG. 131

We divide the strip by this means into domains (I), (II), (III), ... Points of domain (I) correspond to the points of the string at which only disturbances from interior points have arrived, and consequently the infinite parts that we imagine to have been added to the string have no effect here on the motion. At points in domain (II), we already have disturbances coming from the imaginary parts; we shall take, for instance, the point M_0 (x_0, t_0) of domain (II).

Since

$$u(x_0, t_0) = \theta_1(x_0 - at_0) + \theta_2(x_0 + at_0),$$

there are two waves present at this point: firstly the direct wave, from the initial disturbance at the point M_1 of the string with abscissa $x = x_0 - at_0$, and secondly the reverse wave from the point M_2 with abscissa $x = x_0 + at_0$, M_1 being in the present case a real point of the interval $(0, l)$, and M_2 an imagined point. The latter is readily replaced by a real point by noting that, by (28):

$$\theta_2(x_0 + at_0) = \theta_2(l + x_0 + at_0 - l) = -\theta_1(2l - x_0 - at_0),$$

so that the reverse wave $\theta_2(x_0 + at_0)$ is in fact the same as the direct wave $-\theta_1(2l - x_0 - at_0)$ from the point of initial disturbance $M_2'(2l - x_0 - at_0)$ (symmetrical to M_2 about L), which, after arriving at the end L at the instant

$$t = \frac{l - (2l - x_0 - at_0)}{a} = \frac{x_0 + at_0 - l}{a},$$

has reversed its direction and sign and arrived in this form at the instant t_0 at M_0; in other words, *fixing the end $x = l$ has led to a reflection of the displacement wave, corresponding to a change in sign of the displacement whilst the absolute value is retained.*

We find the same phenomenon with waves arriving at the end $x = 0$; we have two waves at points of domain (III): the reverse and the direct, reflected from the end $x = 0$. We obtain at points in domains (IV), (V), (VI), ... waves which have undergone several of these reflections at both ends of the string.

If instead of a boundary condition (25) we had, say, at the end $x = l$, the condition†:

$$\left.\frac{\partial u}{\partial x}\right|_{x=l} = 0\,,$$

we should have, instead of the second of equations (27):

$$\theta_1'(l - at) + \theta_2'(l + at) = 0\,,$$

or, on again replacing at by x:

$$\theta_2'(l + x) = -\,\theta_1'(l - x)\,.$$

Integration of this relationship clearly gives us:

$$\theta_2(l + x) = \theta_1(l - x) + C\,,$$

where C is a constant which we can take as zero without loss of generality, the proof of this being left to the reader. We thus have

$$\theta_2(l + x) = \theta_1(l - x)\,. \tag{29}$$

The physical interpretation of this condition is likewise *reflection at the end $x = l$, though both the sign and magnitude of the displacement are now preserved.*

A remarkably simple example of the use of the above method of characteristics and reflections is given by a "plucked string", which is stretched out at one point at the initial instant with zero initial velocity. The reader may easily verify the method given below for finding the shape of the string at any instant t, the initial shape being assumed given.

† This is met with in the theory of *longitudinal vibrations of rods*, which obey the same differential equation (5) or (6) with different physical values of the constant a. The condition imposed means that the end of the rod is *free*.

OAL gives the initial shape of the string in Fig. 132, the dotted line being the corresponding symmetrical figure with respect to the mid-point of the string $x = l/2$. Let the perpendicular AP to OL be continued to its intersection B' with $A'L$. Let C be the mid-point of AB', which gives us the direction LC.

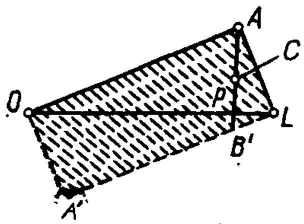

FIG. 132

The shape of the string at any given instant is now obtained by moving a line of intersection parallel to LC from the point A to the point A'; in particular, at the instant $\tau = l/a$, the string takes the dotted polygonal shape $OA'L$.

Figure 133 illustrates the successive forms of the string at the instants:

$$0, \frac{1}{4}\tau, \ \frac{1}{2}\tau, \ \frac{3}{4}\tau, \ \tau.$$

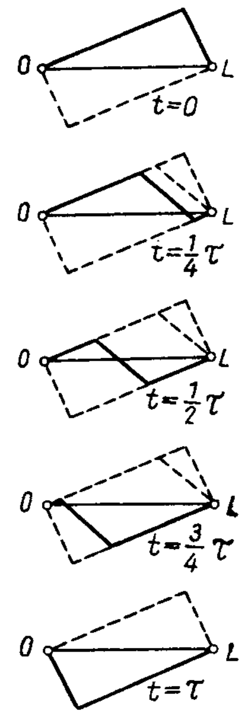

FIG. 133

167. Fourier's method. The transverse vibrations of a string fixed at its ends may also be treated by means of Fourier series. Although this method is not as simple as the above in this particular case, it is worth describing because it can be used in many other problems for which the method of characteristics is inapplicable. We again write down the equations of the problem, in a different order:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, \tag{30}$$

$$u\big|_{x=0} = 0, \qquad u\big|_{x=l} = 0, \tag{31}$$

$$u\bigg|_{t=0} = \varphi(x); \quad \frac{\partial u}{\partial t}\bigg|_{t=0} = \varphi_1(x). \tag{32}$$

Instead of seeking the general solution of equation (30), *we seek a particular solution as the product of two functions, one of which depends only on t and the other only on x:*

$$u = T(t)\, X(x). \tag{33}$$

We have on substituting this in (30):

$$X(x)\, T''(t) = a^2\, T(t)\, X''(x)$$

or

$$\frac{T''(t)}{a^2 T(t)} = \frac{X''(x)}{X(x)}.$$

The function on the left of this equation depends only on t, and that on the right only on x; equality is only possible if both left and right are independent of both t and x, i.e. both sides represent the same constant.

We write this constant as $(-k^2)$:

$$\frac{T''(t)}{a^2 T(t)} = \frac{X''(x)}{X(x)} = -k^2; \tag{34}$$

hence we get two equations:

$$X''(x) + k^2 X(x) = 0; \;\; T''(t) + a^2 k^2 T(t) = 0. \tag{35}$$

The general solutions of these equations are, with $k \neq 0$ [27]:

$$X(x) = C \cos kx + D \sin kx; \;\; T(t) = A \cos akt + B \sin akt, \tag{35_1}$$

where A, B, C, D are arbitrary constants.

We obtain for u, from (33):

$$u = (A \cos akt + B \sin akt)(C \cos kx + D \sin kx). \tag{36}$$

We now choose the constants so that boundary conditions (31) are satisfied, i.e. the factor in x of (36) vanishes for $x = 0$ and $x = l$. This gives:

$$C \cdot 1 + D \cdot 0 = 0; \;\; C \cos kl + D \sin kl = 0.$$

It follows from the first equation that $C = 0$, and the second gives $D \sin kl = 0$. If we take $D = 0$, we have $C = D = 0$ and (36) is identically zero. This is a trivial solution, so that we must take $D \neq 0$ but $\sin kl = 0$.

We thus get an equation giving the parameter k, which has so far remained entirely arbitrary:†

$$\sin kl = 0,$$

i.e.

$$k = \pm \frac{\pi}{l}, \;\; \pm \frac{2\pi}{l}, \ldots, \;\; \pm \frac{n\pi}{l}, \ldots \tag{37}$$

If we substitute $k = n\pi/l$ or $k = -n\pi/l$ in (36), the only difference is in the sign of the sines, and in view of the presence of the arbitrary

† If we had written $(+k^2)$ for the constant in (34) instead of $(-k^2)$, we should have obtained $X(x) = Ce^{kx} + De^{-kx}$ and it would have been quite impossible to satisfy boundary conditions (31).

The same situation is obtained with $k = 0$. Similar remarks apply to later problems for which we use Fourier's method.

constant factors the two solutions will be essentially the same. Hence it is sufficient to take only positive k from values (37). On putting $C = 0$ in (36) and writing A and B for the arbitrary constants AD and BD, we get:

$$u = (A \cos akt + B \sin akt) \sin kx.$$

We still have to substitute one of the values (37) for k. We can take different values for A and B on substituting different values of k. This leads us to an infinite set of solutions:

$$u_n = \left(A_n \cos \frac{n\pi at}{l} + B_n \sin \frac{n\pi at}{l}\right) \sin \frac{n\pi x}{l} \quad (n = 1, 2, \ldots). \tag{38}$$

These solutions satisfy both equation (30) and boundary conditions (31). We now notice that, due to the fact that (30) and (31) are linear and homogeneous equations, if $u_1, u_2, \ldots$ are solutions satisfying these, their sum likewise satisfies the equations (as in the analogous case of ordinary linear homogeneous equations). Hence we have the following solution:

$$u = \sum_{n=1}^{\infty} \left(A_n \cos \frac{n\pi at}{l} + B_n \sin \frac{n\pi at}{l}\right) \sin \frac{n\pi x}{l}. \tag{39}$$

It remains to choose A_n and B_n so as to satisfy also the initial conditions (32). We differentiate solution (39) with respect to t:

$$\frac{\partial u}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{n\pi a}{l} A_n \sin \frac{n\pi at}{l} + \frac{n\pi a}{l} B_n \cos \frac{n\pi at}{l}\right) \sin \frac{n\pi x}{l}. \tag{40}$$

On setting $t = 0$ in (39) and (40), we get by (32):

$$\varphi(x) = \sum_{n=1}^{\infty} A_n \sin \frac{n\pi x}{l}, \quad \varphi_1(x) = \sum_{n=1}^{\infty} \frac{n\pi a}{l} B_n \sin \frac{n\pi x}{l}. \tag{41}$$

The series written represent sine expansions of the given functions $\varphi(x)$ and $\varphi_1(x)$ in the interval $(0, l)$. The coefficients of these expansions are defined by the familiar formulae of [146], which leads us to the following values for A_n and B_n:

$$A_n = \frac{2}{l} \int_0^l \varphi(z) \sin \frac{n\pi z}{l} dz; \quad B_n = \frac{2}{n\pi a} \int_0^l \varphi_1(z) \sin \frac{n\pi z}{l} dz. \tag{42}$$

On substituting these values in (39), we obtain a series which formally meets all our requirements. We give below the sufficient conditions to be imposed on $\varphi(x)$ and $\varphi_1(x)$ so that the sum of the series in fact provides the solution of our problem.

168. Harmonics and standing waves. We bring in the amplitude N_n and initial phase φ_n of a harmonic:

$$A_n \cos\frac{n\pi at}{l} + B_n \sin\frac{n\pi at}{l} = N_n \sin\left(\frac{n\pi at}{l} + \varphi_n\right).$$

Each term of series (39) gives a solution of the problem:

$$\left(A_n \cos\frac{n\pi at}{l} + B_n \sin\frac{n\pi at}{l}\right)\sin\frac{n\pi x}{l} = N_n \sin\left(\frac{n\pi at}{l} + \varphi_n\right)\sin\frac{n\pi x}{l} \quad (43)$$

and represents a so-called *standing wave*, such that points of the string carry out harmonic vibrations of the same phase and with the amplitude

$$N_n \sin\frac{n\pi x}{l},$$

depending on the position of the point. The string emits a sound with this type of vibration; the *pitch* depends on the *frequency* of the vibration:

$$\omega_n = \frac{n\pi a}{l}, \quad (44)$$

whilst the strength depends on the maximum amplitude of the vibration N_n. On assigning the values $1, 2, 3, \ldots$ to n, we get *the fundamental tone of the string and a series of successive overtones,* the frequencies of which or numbers of vibrations per second are proportional to the terms of the natural series of integers $1, 2, 3, \ldots$ The amplitude N_n sin $n\pi x/l$ can be negative for certain values. Its absolute value can be taken on adding π to the phase.

Solution (39), i.e. the emitted sound, is the addition of these separate tones or *harmonics;* their amplitudes, and therefore their influence on the sound, generally diminishes rapidly as the number of the harmonic increases, whilst the total result of their effects produces the *timbre* of the sound, which differs with different musical instruments and is due to the presence of the overtones.

At the points

$$x = 0,\ \frac{l}{n},\ \frac{2l}{n}, \ldots, \frac{(n-1)\,l}{n}, l, \quad (45)$$

the amplitude of the nth harmonic vanishes, since $\sin n\pi x/l = 0$ at these points, which are called *nodes* of the nth harmonic. On the other hand, at

$$x = \frac{l}{2n},\ \frac{3l}{2n}, \ldots, \frac{(2n-1)\,l}{2n} \quad (45_1)$$

the amplitude of the nth harmonic has maxima, since $\sin n\pi x/l$ attains its maximum absolute value at these points, which are known as antinodes of the nth harmonic. With this, the string vibrates as though it were composed of n mutually unconnected separate pieces, each fixed at its boundary nodes. If we press the string exactly at the midpoint, i.e. at the antinode of its fundamental tone, we obtain vanishing not only of this tone but of all the others that have antinodes at this point, i.e. the 3rd, 5th, ... harmonics; whereas the even harmonics, which have nodes at this point, are unaffected, so that the string now emits the octave instead of the fundamental, i.e. the note with twice the number of vibrations per second.

The above method might be distinguished from that of characteristics by calling it the *standing wave method*; but it is generally known as *Fourier's method.*

We can readily show that the solution given by series (39) is completely identical with that obtained above in [166]. We start by recalling the proof in [166] that the use of d'Alembert's formula (16) in the case of a finite string requires the odd continuation in the interval $(-l, 0)$, and thereafter continuation with period $2l$, of the functions $\varphi(x)$ and $\varphi_1(x)$ specified in the interval $(0, l)$. But this method of continuation is exactly equivalent to the sine expansion of these functions [145], i.e. is exactly equivalent to expressions (41) for any x. On substituting these expressions for $\varphi(x)$ and $\varphi_1(x)$ in d'Alembert's formula (17), we in fact arrive at solution (39), as may easily be seen:

$$u=\frac{1}{2}\sum_{n=1}^{\infty} A_n\left[\sin\frac{n\pi(x-at)}{l}+\sin\frac{n\pi(x+at)}{l}\right]+$$
$$+\frac{1}{2a}\int\limits_{x-at}^{x+at}\sum_{n=1}^{\infty}\frac{n\pi a}{l}B_n\sin\frac{n\pi z}{l}\,dz$$

or

$$u=\frac{1}{2}\sum_{n=1}^{\infty} A_n\left[\sin\frac{n\pi(x-at)}{l}+\sin\frac{n\pi(x+at)}{l}\right]+$$
$$+\frac{1}{2}\sum_{n=1}^{\infty} B_n\left[\cos\frac{n\pi(x-at)}{l}-\cos\frac{n\pi(x+at)}{l}\right],$$

whence (39) follows at once.

In the present case, Fourier's method has the disadvantage compared with the characteristics method that series (39) is often very slowly convergent and is unsuitable both for computation and even for the

rigorous proof of the fact that the series is in fact the solution, since this latter involves its differentiation twice term by term, which brings in a factor of n^2 before the nth term. The relationship between the required function and the originally assigned $\varphi(x)$, $\varphi_1(x)$, expressible by series (39), is much more complicated outwardly than the relationship given by the characteristics method. Nevertheless, Fourier's method reveals the extremely important fact that the string has an infinite set of individual harmonics, addition of which gives the total vibration.

On taking into account the discussion of [166], we can say that the sum of series (39) will give the solution of our problem with continuous derivatives up to the second order, if $\varphi(x)$ and $\varphi_1(x)$ possess the properties stated in [166]. If $\varphi(x)$ has continuous derivatives up to the third order and satisfies $\varphi(0) = \varphi''(0) = \varphi(l) = \varphi''(l) = 0$, whilst $\varphi_1(x)$ has continuous derivatives to the second order and $\varphi_1(0) = \varphi_1(l) = 0$, series (39) can be shown to be twice differentiable with respect to x and t. It is also possible to consider solutions of the wave equation with fewer assumptions regarding the initial data, and we shall discuss this in Volume IV. In future applications of Fourier's method, we shall not stipulate the conditions under which the series obtained in fact represent solutions of our problems. The general aspects of Fourier's method are dealt with in Volume IV. The present aim is to indicate the method of solution and the results obtainable. A further point: it follows at once from the arguments of [164] and the characteristics method of [166] that the above solutions are unique, both for an infinite and a finite string. The question of the uniqueness of the solutions of the general wave equation is discussed below.

169. Forced vibrations. We deduced in [163] the equation of the forced vibrations of a string acted on by a force $F(x, t)$ per unit length:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} + f(x, t) \quad \left[f(x, t) = \frac{1}{\varrho} F(x, t)\right]. \tag{46}$$

Boundary and initial conditions must be associated with this equation (taking the case of a string fixed at the ends):

$$u\big|_{x=0} = 0; \; u\big|_{x=l} = 0, \tag{47}$$

$$u\big|_{t=0} = \varphi(x); \; \frac{\partial u}{\partial t}\bigg|_{t=0} = \varphi_1(x). \tag{48}$$

These forced vibrations of a general type can be represented as the result of adding two vibratory motions; one of these is purely forced and due to the action of the force F, the string being assumed still in the state of rest at the initial instant, whereas the other is the free vibration accomplished by the string simply as a result of the initial disturbance without the force being present. This amounts analytically to replacing u by two new functions v and w, in accordance with the formula:

$$u = v + w,$$

where v satisfies the conditions:

$$\frac{\partial^2 v}{\partial t^2} = a^2 \frac{\partial^2 v}{\partial x^2} + f(x\ t), \tag{49}$$

$$v|_{x=0} = 0; \quad v|_{x=l} = 0, \tag{50}$$

$$v|_{t=0} = 0; \ \left.\frac{\partial v}{\partial t}\right|_{t=0} = 0. \tag{51}$$

and provides the purely forced vibration, whilst w satisfies the conditions:

$$\frac{\partial^2 w}{\partial t^2} = a^2 \frac{\partial^2 w}{\partial x^2},$$

$$w|_{x=0} = 0, \qquad w|_{x=l} = 0,$$

$$w|_{t=0} = \varphi(x), \ \left.\frac{\partial w}{\partial t}\right|_{t=0} = \varphi_1(x)$$

and gives the free vibration. We may readily verify, on forming the sum $u = v + w$, that this provides the solution of our problem, i.e. of equations (46), (47), (48).

The methods of finding the free vibrations w have been considered in previous sections, so that we shall deal here only with how to find v. As in the case of free vibrations, we seek v as a series:

$$v(x, t) = \sum_{n=1}^{\infty} T_n(t) \sin \frac{n\pi x}{l}, \tag{52}$$

so that boundary conditions (50) are automatically satisfied, whilst the $T_n(t)$ are obviously different from the functions that we had in [167], since equation (49) is not homogeneous.

We obtain on substituting series (52) in equation (49):

$$\sum_{n=1}^{\infty} T''(t) \sin \frac{n\pi x}{l} = -a^2 \sum_{n=1}^{\infty} T_n(t) \left(\frac{n\pi}{l}\right)^2 \sin \frac{n\pi x}{l} + f(x, t),$$

whence, on replacing $an\pi/l$ by the quantity ω_n (44) [168]:

$$f(x,t) = \sum_{n=1}^{\infty} \left[T_n''(t) + \omega_n^2 T_n(t)\right] \sin\frac{n\pi x}{l}. \tag{53}$$

We can expand $f(x, t)$, considered as a function of x, in a Fourier series in the interval $0 \leqslant x \leqslant l$:

$$f(x,t) = \sum_{n=1}^{\infty} f_n(t) \sin\frac{n\pi x}{l}, \tag{54}$$

the coefficients $f_n(t)$ of this series being dependent on t and given by

$$f_n(t) = \frac{2}{l}\int_0^l f(z,t) \sin\frac{n\pi z}{l}\, dz, \tag{55}$$

On comparing expansions (53) and (54) for the same function $f(x, t)$, we obtain the series of equations

$$T_n''(t) + \omega_n^2 T_n(t) = f_n(t) \qquad (n = 1, 2, \ldots,), \tag{56}$$

defining the functions $T_1(t)$, $T_2(t)$, ...

Having thus defined the functions $T_n(t)$, function (52) now satisfies differential equation (49) and boundary conditions (50). In order to satisfy also the remaining initial conditions (51), we merely have to subject the $T_n(t)$ to these conditions, i.e. we put

$$T_n(0) = 0, \; T_n'(0) = 0, \tag{57}$$

since it is then clear that

$$v\Big|_{t=0} = \sum_{n=1}^{\infty} T_n(0) \sin\frac{n\pi x}{l} = 0; \quad \frac{\partial v}{\partial t}\Big|_{t=0} = \sum_{n=1}^{\infty} T_n'(0) \sin\frac{n\pi x}{l} = 0.$$

The solution of equations (56) and (57) was indicated in [28], whence we readily deduce that:

$$T_n(t) = \frac{1}{\omega_n}\int_0^t f_n(\tau) \sin\omega_n(t-\tau)\, d\tau,$$

or, on substituting expression (55) for $f_n(\tau)$:

$$T_n(t) = \frac{2}{l\omega_n}\int_0^t d\tau \int_0^l f(z,\tau) \sin\omega_n(t-\tau) \sin\frac{n\pi x}{l}\, dz. \tag{58}$$

Substitution of this in (52) gives us the expression for $v(x, t)$. It is easily shown that if $f(x, t)$ has continuous derivatives up to the second

order and $f(0, t) = f(l, t) = 0$, the sum of series (52) is in fact the solution of problem (49)—(51).

We have so far considered non-homogeneity either in the initial conditions (with function w) or in the differential equation (with function v). It is natural to consider also non-homogeneous boundary conditions. If we assume the equation and the initial conditions homogeneous and again let u denote the required function, we get the following problem:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}; \; u\Big|_{x=0} = \omega(t); \; u\Big|_{x=l} = \omega_1(t); \; u\Big|_{t=0} = \frac{\partial u}{\partial t}\Big|_{t=0} = 0.$$

We shall discuss this case of non-homogeneous boundary conditions in Volume IV.

170. Concentrated force. We investigate (58) in the case of a concentrated force at the single point C ($x = c$). We shall denote the magnitude of the force by ϱP instead of by P as in [163]. As has been shown [163], this case can be considered as the limit of the case when the force F acts only on a small interval $(c - \delta, c + \delta)$, F being zero outside this interval, whilst its total magnitude

$$\int_{c-\delta}^{c+\delta} F(z, t)\, dz \to \varrho P(t) \text{ as } \delta \to 0.$$

We have from the second of equations (4):

$$\int_{c-\delta}^{c+\delta} f(z, t)\, dz \to P(t) \text{ as } \delta \to 0.$$

If we note that $f(z, t)$ is zero by hypothesis outside the interval $c - \delta \leqslant z \leqslant c + \delta$ and use the first mean value theorem [I, 95] with the assumption that $f(z, t)$ has an invariable sign in

$$c - \delta \leqslant z \leqslant c + \delta,$$

we get:

$$\int_0^l f(z, t) \sin\frac{n\pi z}{l}\, dz = \int_{c-\delta}^{c+\delta} f(z, t) \sin\frac{n\pi z}{l}\, dz = \sin\frac{n\pi\zeta}{l} \int_{c-\delta}^{c+\delta} f(z, t)\, dz,$$

where ζ is a point of $(c - \delta, c + \delta)$.

In the limit, as $\delta \to 0$:

$$\int_0^l f(z, t) \sin\frac{n\pi z}{l}\, dz \to P(t) \sin\frac{n\pi c}{l},$$

and the function $T_n(t)$, defined as the limit of the right-hand side of (58) as $\delta \to 0$, now becomes

$$T_n(t) = \frac{2}{l\omega_n} \sin \frac{n\pi c}{l} \int_0^t P(\tau) \sin \omega_n (t-\tau)\, d\tau,$$

whilst the forced vibration is given by the expression:

$$v(x,t) = \sum_{n=1}^{\infty} \frac{2}{l\omega_n} \sin \frac{n\pi c}{l} \int_0^t P(\tau) \sin \omega_n (t-\tau)\, d\tau \cdot \sin \frac{n\pi x}{l}. \tag{59}$$

This formula shows that certain overtones can be absent in forced vibrations, these being the overtones for which

$$\sin \frac{n\pi c}{l} = 0,$$

i.e. which have a node at the point C of application of the force.

We shall dwell on the case when the impressed force is of a harmonic oscillatory type and we have to put

$$P(t) = P_0 \sin(\omega t + \varphi_0),$$

or, if we take the phase $\varphi_0 = 0$ for simplicity:

$$P(t) = P_0 \sin \omega t.$$

The expression for $T_n(t)$ now gives:

$$T_n(t) = \frac{P_0}{l\omega_n} \sin \frac{n\pi c}{l} \int_0^t 2 \sin \omega\tau \sin \omega_n (t-\tau)\, d\tau =$$

$$= -\frac{P_0}{l\omega_n} \sin \frac{n\pi c}{l} \int_0^t \left\{\cos\left[\omega_n t - (\omega_n - \omega)\tau\right] - \cos\left[\omega_n t - (\omega_n + \omega)\tau\right]\right\} d\tau =$$

$$= \frac{-2\omega P_0}{l\omega_n(\omega_n^2 - \omega^2)} \sin \frac{n\pi c}{l} \sin \omega_n t + \frac{2P_0}{l(\omega_n^2 - \omega^2)} \sin \frac{n\pi c}{l} \sin \omega t.$$

If the frequency of the impressed force is not coincident with any of the proper frequencies ω_n, none of the denominators $(\omega_n^2 - \omega^2)$ vanishes; where as if ω approaches one of the ω_n, the corresponding denominator decreases and the particular $T_n(t)$ becomes very large compared with the rest, i.e. the phenomenon of resonance takes place. Finally, if $\omega = \omega_n$, the above expression for $T_n(t)$ becomes meaningless and has to be replaced by another.

Substitution of the expression obtained for the $T_n(t)$ in (52) gives us:

$$v(x,t) = \frac{-2\omega P_0}{l} \sum_{n=1}^{\infty} \frac{1}{\omega_n} \frac{\sin \frac{n\pi c}{l}}{\omega_n^2 - \omega^2} \sin \omega_n t \sin \frac{n\pi x}{l} +$$

$$+ \frac{2P_0}{l} \sin \omega t \sum_{n=1}^{\infty} \frac{\sin \frac{n\pi c}{l}}{\omega_n^2 - \omega^2} \sin \frac{n\pi x}{l}.$$

The first term on the right has the form of the proper vibrations, whilst the second has the same frequency as the disturbing force. We shall neglect the first term with free vibrations $w(x, t)$ and only concern ourselves with the second, writing it as $V(x, t)$:

$$V(x, t) = \frac{2P_0}{l} \sin \omega t \sum_{n=1}^{\infty} \frac{\sin \frac{n\pi c}{l}}{\omega_n^2 - \omega^2} \sin \frac{n\pi x}{l},$$

or, on setting $\alpha^2 = \omega^2 l^2/a^2 \pi^2$:

$$V(x, t) = \frac{2P_0 l}{a^2 \pi^2} \sin \omega t \sum_{n=1}^{\infty} \frac{\sin \frac{n\pi c}{l}}{n^2 - \alpha^2} \sin \frac{n\pi x}{l}. \tag{60}$$

The sum:

$$\sum_{n=1}^{\infty} \frac{\sin \frac{n\pi c}{l}}{n^2 - \alpha^2} \sin \frac{n\pi x}{l}$$

can be evaluated by the method indicated in [159]. Instead of proceeding along these lines, however, we shall approach the problem in a new way by considering the concentrated force directly instead of as the limiting case of a continuously distributed force.

The point C of application of the force divides the string into two parts $(0, c)$ and (c, l). We consider these two parts individually, denoting the ordinate of the first part by $u_1(x, t)$ and the ordinate of the second by $u_2(x, t)$. We get the following equations for the two functions u_1 and u_2:

$$\frac{\partial^2 u_1}{\partial t^2} = a^2 \frac{\partial^2 u_1}{\partial x^2} \text{ for } 0 < x < c, \tag{61}$$

$$\frac{\partial^2 u_2}{\partial t^2} = a^2 \frac{\partial^2 u_2}{\partial x^2} \text{ for } c < x < l, \tag{61$_1$}$$

since there are no external forces inside the intervals $(0, c)$ and (c, l). Further, we have the conditions for the fixed ends:

$$u_1 |_{x=0} = 0, \; u_2 |_{x=l} = 0, \tag{62}$$

hen the condition for continuity of the string at $x = c$:

$$u_1 |_{x=c} = u_2 |_{x=c}, \tag{63}$$

and finally, the condition for equilibrium of the forces acting at $x = c$ [163]:

$$\left.\frac{\partial u_2}{\partial x}\right|_{x=c} - \left.\frac{\partial u_1}{\partial x}\right|_{x=c} = -\frac{\varrho}{T_0} P(t) = -\frac{1}{a^2} P(t) \dagger \tag{64}$$

We confine ourselves to the case of a harmonic force

$$P(t) = P_0 \sin \omega t$$

† With our present notation, we have to replace P in equation (7) [163] by $\varrho P(t)$, and $(\partial u/\partial x)_+$, $(\partial u/\partial x)_-$ by $\partial u_2/\partial x$, $\partial u_1/\partial x$.

and distinguish in the vibrations produced those of the same frequency ω We shall look for these vibrations in the form:

$$u(x, t) = X(x) \sin \omega t,$$

where, however, we must have different expressions for the function $X(x)$ in the intervals $(0, c)$ and (c, l); as a result of this, we put:

$$u_1 = X_1(x) \sin \omega t, \; u_2 = X_2(x) \sin \omega t. \tag{65}$$

On substituting in equations (61) and (61_1), we have:

$$-\omega^2 \sin \omega t \, X_1(x) = a^2 X_1''(x) \sin \omega t,$$

whence

$$X_1''(x) + \frac{\omega^2}{a^2} X_1(x) = 0,$$

and similarly,

$$X_2''(x) + \frac{\omega^2}{a^2} X_2(x) = 0.$$

We thus obtain, using [27]:

$$X_1(x) = C_1' \cos \frac{\omega}{a} x + C_2' \sin \frac{\omega}{a} x; \; X_2(x) = C_1'' \cos \frac{\omega}{a} x + C_2'' \sin \frac{\omega}{a} x.$$

Conditions (62) give us:

$$C_1' = 0, \; C_1'' \cos \frac{\omega l}{a} + C_2'' \sin \frac{\omega l}{a} = 0,$$

whence it follows that we can put

$$C_1'' = C_2 \sin \frac{\omega l}{a}, \; C_2'' = -C_2 \cos \frac{\omega l}{a},$$

where C_2 is an arbitrary constant. On denoting the arbitrary constant C_2' by C_1 for the sake of symmetry, we get:

$$X_1(x) = C_1 \sin \frac{\omega x}{a}, \; X_2(x) = C_2 \sin \frac{\omega(l-x)}{a}.$$

The continuity condition (63) now gives:

$$C_1 \sin \frac{\omega c}{a} \sin \omega t = C_2 \sin \frac{\omega(l-c)}{a} \sin \omega t.$$

It only remains to satisfy the last condition (64), from which we obtain

$$-\frac{\omega}{a} C_2 \cos \frac{\omega(l-c)}{a} \sin \omega t - \frac{\omega}{a} C_1 \cos \frac{\omega c}{a} \sin \omega t = -\frac{P_0}{a^2} \sin \omega t.$$

Hence constants C_1 and C_2 are determined by the system of equations:

$$C_1 \sin \frac{\omega c}{a} - C_2 \sin \frac{\omega(l-c)}{a} = 0; \quad C_1 \cos \frac{\omega c}{a} + C_2 \cos \frac{\omega(l-c)}{a} = \frac{P_0}{a\omega},$$

whence we obtain by simple working:

$$C_1 = \frac{P_0}{a\omega} \frac{\sin \frac{\omega (l-c)}{a}}{\sin \frac{\omega l}{a}}, \quad C_2 = \frac{P_0}{a\omega} \frac{\sin \frac{\omega c}{a}}{\sin \frac{\omega l}{a}},$$

the final solution of the problem being given by expressions (65) in the form:

$$u(x, t) = \begin{cases} \dfrac{P_0}{a\omega} \dfrac{\sin \frac{\omega (l-c)}{a}}{\sin \frac{\omega l}{a}} \sin \dfrac{\omega x}{a} \sin \omega t \quad \text{for} \quad 0 < x < c, \\ \\ \dfrac{P_0}{a\omega} \dfrac{\sin \frac{\omega c}{a}}{\sin \frac{\omega l}{a}} \sin \dfrac{\omega (l-x)}{a} \sin \omega t \quad \text{for} \quad c < x < l. \end{cases} \tag{66}$$

The reader may easily verify that solutions (66) and (60) for $V(x, t)$ are identical, by expanding (66) as a Fourier sine series.

171. Poisson's formula. By analogy with an infinite string, we now consider the solution of the general wave equation:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) \tag{67}$$

in unbounded space with given initial conditions. We start by deducing an auxiliary proposition. We shall find it more convenient to write the coordinates (x, y, z) as (x_1, x_2, x_3). Let $\omega(x_1, x_2, x_3)$ be any continuous function with continuous derivatives up to the second order in a domain D or throughout space. All future arguments will be in reference to this domain. We consider the values of the function ω on the surface $C_r(x_1, x_2, x_3)$ of a sphere with centre at the point (x_1, x_2, x_3) and radius r. The coordinates of points of the sphere can be written as:

$$\xi_1 = x_1 + \alpha_1 r; \;\; \xi_2 = x_2 + \alpha_2 r; \;\; \xi_3 = x_3 + \alpha_3 r,$$

where $(\alpha_1, \alpha_2, \alpha_3)$ are the direction-cosines of radii of the sphere. We can write these latter in the form:

$$\alpha_1 = \sin\theta \cos\varphi; \;\; \alpha_2 = \sin\theta \sin\varphi; \;\; \alpha_3 = \cos\theta,$$

where the angle θ varies from 0 to π and the angle φ from 0 to 2π. We let $\mathrm{d}_1\sigma$ denote an elementary area of the unit sphere and $\mathrm{d}_r\sigma$ an elementary area of the sphere of radius r:

$$\mathrm{d}_1\sigma = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi; \;\; \mathrm{d}_r\sigma = r^2 \, \mathrm{d}_1\sigma = r^2 \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi.$$

We consider the arithmetic mean of the values of the function ω over the spherical surface $C_r(x_1, x_2, x_3)$, i.e. the integral of $\omega(x_1, x_2, x_3)$ over the spherical surface divided by the surface-area. The value of the integral here obviously depends on the choice of the centre (x_1, x_2, x_3) and the radius r of the sphere, and the arithmetic mean will be a function of the four variables (x_1, x_2, x_3, r). We can write the arithmetic mean either as

$$v(x_1, x_2, x_3, r) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \omega(x_1 + \alpha_1 r;\ x_2 + \alpha_2 r;\ x_3 + \alpha_3 r)\, d_1\sigma, \tag{68}$$

or as

$$v(x_1, x_2, x_3, r) = \frac{1}{4\pi r^2} \iint_{C_r} \omega(x_1 + \alpha_1 r;\ x_2 + \alpha_2 r;\ x_3 + \alpha_3 r)\, d_r\sigma.$$

We shall prove that, for any choice of function ω, the function v satisfies the same partial differential equation, viz:

$$\frac{\partial^2 v}{\partial r^2} - \Delta v + \frac{2}{r} v_r = 0, \tag{69}$$

where as usual,

$$\Delta v = \frac{\partial^2 v}{\partial x_1^2} + \frac{\partial^2 v}{\partial x_2^2} + \frac{\partial^2 v}{\partial x_3^2}.$$

The integration in (68) is carried out over the surface of the unit sphere, and we can differentiate with respect to x_i under the integral sign. We thus have:

$$\Delta v = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \Delta\omega(x_i + \alpha_i r)\, d_1\sigma$$

and

$$\frac{\partial v}{\partial r} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \sum_{k=1}^{3} \frac{\partial \omega}{\partial x_k} \alpha_k\, d_1\sigma.$$

We can transform the last integral to a surface integral over the sphere $C_r(x_1, x_2, x_3)$:

$$\frac{\partial v}{\partial r} = \frac{1}{4\pi r^2} \iint_{C_r} \sum_{k=1}^{3} \frac{\partial \omega}{\partial x_k} \alpha_k\, d_r\sigma,$$

and application of Ostrogradskii's formula gives us

$$\frac{\partial v}{\partial r} = \frac{1}{4\pi r^2} \iiint_{D_r} \Delta\omega\, dv, \tag{70}$$

where D_r is the sphere with centre (x_1, x_2, x_3) and radius r. The last expression is the product of two functions of r: the fraction $1/4\pi r^2$ and the integral. The derivative with respect to r of the triple integral over the sphere D_r is equal to the integral of the same integrand over the surface C_r of this sphere. To see this, we need only say write the integral over D_r in spherical coordinates. Hence a second differentiation with respect to r gives us:

$$\frac{\partial^2 v}{\partial r^2} = -\frac{1}{2\pi r^3}\iiint\limits_{D_r} \Delta\omega \, \mathrm{d}v + \frac{1}{4\pi r^2}\iint\limits_{C_r} \Delta\omega \, \mathrm{d}_r\sigma .$$

If we substitute the above expressions for the different derivatives in equation (69), we see once that the equation is in fact satisfied. If $r \to 0$, it immediately follows from (68) that $v(x_1, x_2, x_3)$ tends to $\omega(x_1, x_2, x_3)$, whilst it follows from (70) that $\partial v/\partial r$ tends to zero, since by the mean value theorem, the triple integral in (70) is of the order r^3, whilst we have r^2 in the denominator. We thus arrive at the following theorem:

THEOREM. *For any choice of a function ω having continuous derivatives up to the second order, the function v defined by equation* (68) *satisfies equation* (69) *and the initial conditions:*

$$v\Big|_{r=0} = \omega(x_1, x_2, x_3); \quad \frac{\partial v}{\partial r}\Big|_{r=0} = 0. \tag{71}$$

We use this theorem to prove that the function

$$u(x_1, x_2, x_3, t) = tv(x_1, x_2, x_3, at) \tag{72}$$

satisfies the wave equation

$$\frac{\partial^2 u}{\partial t^2} = a^2\left(\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2}\right) \tag{73}$$

and the initial conditions

$$u\Big|_{t=0} = 0; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \omega(x_1, x_2, x_3). \tag{74}$$

We have in fact:

$$\frac{\partial u}{\partial t} = v(x_1, x_2, x_3, at) + at\,\frac{\partial v(x_1, x_2, x_3, at)}{\partial r},$$

$$\frac{\partial^2 u}{\partial t^2} = 2a\,\frac{\partial v(x_1, x_2, x_3, at)}{\partial r} + a^2 t\,\frac{\partial^2 v(x_1, x_2, x_3, at)}{\partial r^2},$$

$$\Delta u = t\,\Delta v(x_1, x_2, x_3, at),$$

where e.g. $\partial v(x_1, x_2, x_3, at)/\partial r$ is the value of the derivative $\partial v(x_1, x_2, x_3, r)/\partial r$ at $r = at$. On substituting the above expressions in equation (73), we obtain equation (69) for v with $r = at$, which is in fact valid as shown above. Initial conditions (74) are obtained immediately from (71). Inasmuch as (73) is a linear homogeneous equation with constant coefficients, we can say that the function $u_1 = \partial u/\partial t$ also satisfies the equation. We find the initial conditions for this function with $t = 0$. Initial conditions (74) give us at once:

$$u_1\big|_{t=0} = \omega(x_1, x_2, x_3).$$

We have for the derivative $\partial u_1/\partial t = \partial^2 u/\partial t^2$, by (73):

$$\frac{\partial u_1}{\partial t}\bigg|_{t=0} = a^2\left(\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2}\right)\bigg|_{t=0},$$

whence we find on differentiating the first of initial conditions (74) with respect to the coordinates:

$$\frac{\partial u_1}{\partial t}\bigg|_{t=0} = 0.$$

Thus the derivative with respect to t of the above solution satisfying initial conditions (74) of wave equation (73) is likewise a solution of the wave equation and satisfies the initial conditions:

$$u_1\bigg|_{t=0} = \omega(x_1, x_2, x_3); \quad \frac{\partial u_1}{\partial t}\bigg|_{t=0} = 0. \tag{74_1}$$

If, on returning to the previous coordinate notation, we take some function $\varphi_1(x, y, z)$ for $\omega(x, y, z)$ in the first case of initial conditions (74), and some other function $\varphi(x, y, z)$ for $\omega(x, y, z)$ in the second case of initial conditions (74_1), and add the solutions thus found, we arrive at a solution of equation (67) satisfying the initial conditions:

$$u\bigg|_{t=0} = \varphi(x, y, z); \quad \frac{\partial u}{\partial t}\bigg|_{t=0} = \varphi_1(x, y, z). \tag{75}$$

If we write briefly $T_r\{\omega(M)\}$ for the arithmetic mean of the function ω over the sphere with centre $M(x, y, z)$ and radius r, by the above arguments, we can write the solution of equation (67) satisfying initial conditions (75) in the form:

$$u(M, t) = tT_{at}\{\varphi_1(M)\} + \frac{\partial}{\partial t}\left[tT_{at}\{\varphi(M)\}\right]. \tag{76}$$

This equation is generally known as *Poisson's formula*. It can evidently be written in the form

$$u(x, y, z, t) = \frac{t}{4\pi}\int_0^{2\pi}\int_0^{\pi} \varphi_1(\alpha, \beta, \gamma)\, d_1\sigma + \frac{\partial}{\partial t}\left[\frac{t}{4\pi}\int_0^{2\pi}\int_0^{\pi} \varphi(\alpha, \beta, \gamma)\, d_1\sigma\right], \qquad (76_1)$$

where $d_1\sigma = \sin\theta \, d\theta \, d\varphi$ and (α, β, γ) are the coordinates of a variable point on the sphere:

$$\alpha = x + at\sin\theta\cos\varphi, \quad \beta = y + at\sin\theta\sin\varphi; \quad \gamma = z + at\cos\theta. \qquad (77)$$

The above arguments show that the function u defined by formula (76) in fact satisfies equation (67) and conditions (75) if $\varphi_1(x, y, z)$ has continuous derivatives up to the second order and $\varphi(x, y, z)$ up to the third order. The last statement is bound up with the fact that the second term in (76) contains a differentiation with respect to t.

If $\varphi(x, y, z)$ and $\varphi_1(x, y, z)$ have less satisfactory differential properties, as happens, for instance, in problems with concentrated initial disturbances, it seems natural to suppose that formula (76_1) still gives the solution. But in this case, the solution is in fact generalized and not classical (see Vol. IV).

We shall see later that the problem treated can only have one solution.

Let us suppose that the initial disturbance is concentrated in a bounded volume (v) with surface (σ), i.e. $\varphi(N)$ and $\varphi_1(N)$ are zero outside (v), and let the point M be situated outside (v). With $t < d/a$, where d is the shortest distance from M to (σ), the sphere (S_{at}) is situated outside (v), both the above functions are zero on (S_{at}), and (76) gives $u(M, t) = 0$, i.e. rest at the point M. At the instant $t = d/a$, the surface (S_{at}) touches (σ) and the forward wave-front passes through M. Finally, with $t > D/a$, where D is the greatest distance from M to a point of (σ), the sphere (S_{at}) will again be situated outside (v) [volume (v) will be entirely inside (S_{at})], and (76) again gives $u(M, t) = 0$. The instant $t = D/a$ corresponds to the passage of the rear wave-front through M, after which $u(M, t)$ vanishes at this point instead of becoming a constant as in the case of a string (i.e. for a plane wave). The forward wave-front at a given instant t consists of the surface which separates points already vibrating from points which have not yet begun to vibrate. It follows from the above that every point of this surface has a shortest distance at from (σ). The surface is readily shown to be the envelope of the family of spheres of radius at with centres on the surface (σ). As we shall see, the constant a is the speed of propagation of the wave-front.

172. Cylindrical waves. We refer space to Cartesian coordinates and assume that functions $\varphi(x, y, z)$ and $\varphi_1(x, y, z)$ depend only on x and y, i.e. they preserve a constant value along any straight line parallel to axis OZ. If the point $M(x, y, z)$ is displaced parallel to OZ, it is clear that the right-hand side of (76_1) remains unchanged, i.e. the function $u(x, y, z, t)$ is also independent of z, and (76_1) gives us the solution of the equation:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \tag{78}$$

with the initial conditions:

$$u\Big|_{t=0} = \varphi(x, y)\,; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \varphi_1(x, y). \tag{79}$$

We can consider this solution whilst remaining exclusively on the xy plane. For this, we have to transform the integral over a sphere in (76_1) to an integral over a circle in the xy plane. We take a point $M(x, y)$ on the xy plane. The points with coordinates (α, β, γ) defined in accordance with (77) with $z = 0$ are variable points of the sphere (S_{at}) with centre $M(x, y, 0)$ and radius at. An elementary surface area of this sphere is given by $\mathrm{d}S_{at} = a^2 t^2 \,\mathrm{d}_1\sigma$. Parts of the sphere situated above and below the xy plane project on to this plane as circles (C_{at}) with centre M and radius at. The elementary area of the projection $\mathrm{d}C_{at}$ is connected with the elementary surface area of the sphere $\mathrm{d}S_{at}$ by the formula [62]:

$$\mathrm{d}S_{at} = \frac{\mathrm{d}C_{at}}{\cos(n, Z)},$$

where n is the direction of the normal to (S_{at}), i.e. the radius of the sphere, forming an acute angle with the z axis. If N is a variable point of the sphere and N_1 its projection on the xy plane, it is clear from elementary geometry that

$$\cos(n, Z) = \frac{\overline{NN_1}}{\overline{MN}} = \frac{\sqrt{a^2t^2 - (\alpha - x)^2 - (\beta - y)^2}}{at},$$

where (α, β) are the coordinates of a variable point of the circle (C_{at}). On making all these substitutions in the first integral of (76_1) and noticing that circle (C_{at}) is obtained both from the upper and the

lower parts of sphere (S_{at}), we get the following transformation for the first integral:

$$\frac{t}{4\pi}\int_0^{2\pi}\int_0^{\pi}\varphi_1(\alpha,\beta,\gamma)\,\mathrm{d}_1\sigma = \frac{1}{4\pi a^2 t}\iint_{(S_{at})}\varphi_1(\alpha,\beta)\,\mathrm{d}S_{at} =$$

$$= \frac{1}{2\pi a}\iint_{(C_{at})}\frac{\varphi_1(\alpha,\beta)}{\sqrt{a^2t^2-(\alpha-x)^2-(\beta-y)^2}}\,\mathrm{d}C_{at}.$$

On applying the same transformation to the second integral and writing the elementary area $\mathrm{d}C_{at}$ on the xy plane in the form $\mathrm{d}\alpha\,\mathrm{d}\beta$, we finally get the following formula for the function satisfying equation (78) and conditions (79):

$$u(x,y,t) = \frac{1}{2\pi a}\iint_{(C_{at})}\frac{\varphi_1(\alpha,\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta}{\sqrt{a^2t^2-(\alpha-x)^2-(\alpha-y)^2}} + $$

$$+ \frac{\partial}{\partial t}\left[\frac{1}{2\pi a}\iint_{(C_{at})}\frac{\varphi(\alpha,\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta}{\sqrt{a^2t^2-(\alpha-x)^2-(\beta-y)^2}}\right]. \tag{80}$$

Let the initial disturbance be confined to a finite domain (B) of the xy plane with contour (l), i.e. $\varphi(x, y)$ and $\varphi_1(x, y)$ are zero outside (B). For instants $t < d/a$, where d is the shortest distance from M to contour (l), the circle (C_{at}) has no common points with (B), $\varphi(x, y)$ and $\varphi_1(x, y)$ are zero throughout (C_{at}), and (80) gives $u(x, y, t) = 0$. At the instant $t = d/a$ the forward wave-front reaches M. For $t >$ $> D/a$, where D is the greatest distance from M to contour (l), domain (B) lies wholly inside the circle (C_{at}) and the integration in (80) must be carried out simply over (B) since $\varphi(x, y)$ and $\varphi_1(x, y)$ vanish outside (B), i.e.

$$u(x,y,t) = \frac{1}{2\pi a}\iint_{(B)}\frac{\varphi_1(\alpha,\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta}{\sqrt{a^2t^2-(\alpha-x)^2-(\beta-y)^2}} + $$

$$+ \frac{\partial}{\partial t}\left[\frac{1}{2\pi a}\iint_{(B)}\frac{\varphi(\alpha,\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta}{\sqrt{a^2t^2-(\alpha-x)^2-(\beta-y)^2}}\right].$$

In the present case, the function $u(x, y, t)$ does not vanish after passage of the rear wave-front at the instant $t = D/a$, as in the case of three-dimensional space, nor does it become constant as in the case of a string. But in view of the presence of $a^2 t^2$ in the denominator we can still affirm that $u(x, y, t)$ tends to zero on indefinite increase of t.

The phenomenon occurring here is known as wave diffusion after passage of the rear front. All our arguments have been carried through

whilst remaining on the xy plane. In three-dimensional space, equation (7⁸) corresponds to so-called cylindrical waves.

173. The case of n-dimensional space. The results obtained in [171] can be generalized immediately for the case of any number of dimensions. Let us consider n-dimensional space with coordinates $(x_1, x_2, \ldots, x_n)$. The volume of a sphere of radius r in such space is given by the formulae [99]:

$$v_n(r) = \frac{(2\pi)^{\frac{1}{2}n}}{2 \cdot 4 \cdot 6 \ldots (n-2) \cdot n} r^n \qquad (n \quad \text{even}),$$

$$v_n(r) = \frac{2^{\frac{1}{2}(n+1)} \pi^{\frac{1}{2}(n-1)}}{1 \cdot 3 \cdot 5 \ldots (n-2)\, n} r^n \qquad (n \quad \text{odd}).$$

On differentiating these expressions with respect to r, we get the surface area of the sphere:

$$\sigma_n(r) = \frac{(2\pi)^{\frac{1}{2}n}}{2 \cdot 4 \cdot 6 \ldots (n-2)} r^{n-1} \qquad (n \quad \text{even}),$$

$$\sigma_n(r) = \frac{2^{\frac{1}{2}(n+1)} \pi^{\frac{1}{2}(n-1)}}{1 \cdot 3 \cdot 5 \ldots (n-2)} r^{n-1} \qquad (n \quad \text{odd}).$$

The direction-cosines α_k of radii of the sphere are given in terms of $(n-1)$ angles by the formulae:

$$\begin{aligned}
\alpha_1 &= \cos\theta_1 \\
\alpha_2 &= \sin\theta_1 \cos\theta_2 \\
\alpha_3 &= \sin\theta_1 \sin\theta_2 \cos\theta_3 \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
\alpha_{n-2} &= \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-3} \cos\theta_{n-2} \\
\alpha_{n-1} &= \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-2} \cos\varphi \\
\alpha_n &= \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-2} \sin\varphi,
\end{aligned}$$

where

$$0 \leqslant \theta_k \leqslant \pi\,; \quad 0 \leqslant \varphi < 2\pi.$$

An elementary surface area of the unit sphere becomes

$$d_1\sigma = \sin^{n-2}\theta_1 \sin^{n-3}\theta_2 \ldots \sin\theta_{n-2}\, d\theta_1\, d\theta_2 \ldots d\theta_{n-2}\, d\varphi\,,$$

and of a sphere of radius r:

$$d_r\sigma = r^{n-1}\, d_1\sigma.$$

Let ω be a function with continuous derivatives up to the second order given in the space R_n. Its arithmetic mean over a sphere of radius r and centre at $(x_1, \ldots, x_n)$ is given by:

$$v(x_1, x_2, \ldots, x_n, r) = \frac{1}{\sigma_n(1)} \int \ldots \int \omega(x_1 + \alpha_1 r,\ x_2 + \alpha_2 r, \ldots, x_n + \alpha_n r)\, d_1\sigma$$

or

$$v(x_1, x_2, \ldots, x_n, r) = \frac{1}{\sigma_n(r)} \int \cdots \int \omega(x_1 + a_1 r,\ x_2 + a_2 r, \ldots, x_n + a_n r)\, d_r\sigma .$$

We can show, precisely as above, that the function v satisfies the differential equation:

$$\frac{\partial^2 v}{\partial r^2} - \Delta v + \frac{n-1}{r}\frac{\partial v}{\partial r} = 0$$

and the initial conditions:

$$v\Big|_{r=0} = \omega(x_1, \ldots, x_n)\,; \quad \frac{\partial v}{\partial r}\Big|_{r=0} = 0 .$$

The results obtained can be used to derive definitive formulae for the wave equation with any number of independent variables. We shall simply state final results for the general case.

The solution of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = a^2\left(\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}\right) \tag{81}$$

with initial conditions

$$u\Big|_{t=0} = 0\,; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \omega(x_1, x_2, \ldots, x_n)$$

has the form, with n odd:

$$u(x_1, \ldots, x_n, t) = \frac{2^{\frac{n-3}{2}}}{1 \cdot 3 \ldots (n-2)} \cdot \frac{\partial^{\frac{n-3}{2}}}{\partial (t^2)^{\frac{n-3}{2}}}\left[t^{n-2}\, T_{at}\{\omega(x_i)\}\right] \tag{82$_1$}$$

and with n even:

$$u(x_1, \ldots, x_n, t) =$$

$$= \frac{2^{\frac{n-2}{2}}}{2 \cdot 4 \ldots (n-2)} \frac{\partial}{\partial t}\int_0^{at} \frac{r}{\sqrt{t^2 - r^2}} \frac{\partial^{\frac{n-2}{2}}}{\partial (r^2)^{\frac{n-2}{2}}}\left[r^{n-2}\, T_r\{\omega(x_i)\}\right] dr, \tag{82$_2$}$$

where $T_\varrho\{\omega(x_i)\}$ is the arithmetic mean of the function $\omega(x_1, x_2, \ldots, x_n)$ over a sphere of radius ϱ and centre at $(x_1, x_2, \ldots, x_n)$. A sufficient requirement for verifying (82_1) and (82_2) is that $\omega(x_1, x_2, \ldots, x_n)$ should have continuous derivatives up to order $(n+1)/2$ in the case of n odd, and up to $(n+2)/2$ in the case of even n.

174. Non-homogeneous wave equation. We take the non-homogeneous wave equation:

$$\frac{\partial^2 u}{\partial t^2} = a^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + f(x, y, z, t) \tag{83}$$

in infinite space and look for a solution of it satisfying the zero initial conditions:

$$u\Big|_{t=0} = 0;\ \frac{\partial u}{\partial t}\Big|_{t=0} = 0. \tag{84}$$

Addition of this solution to the solution of the homogeneous equation satisfying initial conditions (75) will give us a solution of equation (83) satisfying initial conditions (75).

In order to solve the above problem, we bring in the solution of the homogeneous equation

$$\frac{\partial^2 w}{\partial t^2} = a^2\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right), \tag{85}$$

satisfying the initial conditions

$$w\Big|_{t=\tau} = 0;\ \frac{\partial w}{\partial t}\Big|_{t=\tau} = f(x, y, z, \tau), \tag{86}$$

$t = \tau$ being taken as the initial instant instead of $t = 0$, where τ is a parameter. The function w is given by Poisson's formula, except that t has to be replaced here by $(t - \tau)$, on account of our new initial instant. We thus have:

$$w(x, y, z, t; \tau) =$$
$$= \frac{t-\tau}{4\pi}\int_0^{2\pi}\int_0^{\pi} f\left[x + \alpha_1 a(t-\tau),\ y + \alpha_2 a(t-\tau),\ z + \alpha_3 a(t-\tau), \tau\right] \mathrm{d}_1\sigma, \tag{87}$$

where

$$\alpha_1 = \sin\theta\cos\varphi;\ \alpha_2 = \sin\theta\sin\varphi;\ \alpha_3 = \cos\theta. \tag{88}$$

We notice that w depends on the parameter τ as well as on the ordinary independent variables (x, y, z, t). We now define a function $u(x, y, z, t)$ by the formula:

$$u(x, y, z, t) = \int_0^t w(x, y, z, t; \tau)\,\mathrm{d}\tau \tag{89}$$

and show that it satisfies non-homogeneous equation (83) and zero initial conditions (84). We have:

$$\frac{\partial u}{\partial t} = \int_0^t \frac{\partial w(x, y, z, t; \tau)}{\partial t}\,\mathrm{d}\tau + w(x, y, z, t; \tau)\Big|_{\tau=t}. \tag{90}$$

The term outside the integral is zero by the first of conditions (86). A second differentiation with respect to t gives us:

$$\frac{\partial^2 u}{\partial t^2} = \int_0^t \frac{\partial^2 w(x, y, z, t; \tau)}{\partial t^2} \mathrm{d}\tau + \left.\frac{\partial w(x, y, z, t; \tau)}{\partial t}\right|_{\tau=t},$$

the term outside the integral being in this case equal to $f(x, y, z, t)$ by the second of conditions (86), i.e.

$$\frac{\partial^2 u}{\partial t^2} = \int_0^t \frac{\partial^2 w(x, y, z, t; \tau)}{\partial t^2} \mathrm{d}\tau + f(x, y, z, t).$$

Differentiation of expression (89) with respect to the coordinates simply requires differentiation of the integrand:

$$\Delta u = \int_0^t \Delta w(x, y, z, t; \tau)\, \mathrm{d}\tau.$$

It follows at once from the last two expressions and equation (85) that u satisfies equation (83). Initial conditions (84) follow immediately from (89) and (90) on taking account of the fact that the term outside the integral in (90) is zero as shown above. Expression (89) thus gives the solution of equation (83) with initial conditions (84). On substituting for $w(x, y, z, t; \tau)$ in (89) in accordance with (87), we get:

$$u(x, y, z, t) = \frac{1}{4\pi}\int_0^t (t-\tau)\left[\int_0^{2\pi}\int_0^{\pi} f\left[x + a_1 a(t-\tau),\ \ y + a_2 a(t-\tau),\right.\right.$$

$$\left.\left. z + a_3 a(t-\tau),\ \ \tau\right] \mathrm{d}_1\sigma\right] \mathrm{d}\tau.$$

This expression for u can be written in a new form. Instead of τ, we introduce a new variable of integration: $r = a(t - \tau)$. On carrying out the change of variables, we get:

$$u(x, y, z, t) =$$

$$= \frac{1}{4\pi a^2}\int_0^{at}\int_0^{2\pi}\int_0^{\pi} f\left(x + a_1 r,\ y + a_2 r,\ z + a_3 r,\ t - \frac{r}{a}\right) r^2 \sin\theta\, \mathrm{d}r\, \mathrm{d}\theta\, \mathrm{d}\varphi,$$

or, on multiplying and dividing by r:

$$u(x, y, z, t) =$$

$$= \frac{1}{4\pi a^2} \int\limits_0^{at} \int\limits_0^{2\pi} \int\limits_0^{\pi} \frac{f\left(x + a_1 r,\, y + a_2 r,\, z + a_3 r,\, t - \dfrac{r}{a}\right)}{r} r^2 \sin\theta \, dr \, d\theta \, d\varphi.$$

On taking account of expressions (88) for the a_k and recalling the expression for an elementary volume in spherical coordinates, we see that the three quadratures in the above formula are equivalent to the triple integral over the sphere D_{at} with centre (x, y, z) and radius at. On bringing in the variable point

$$\xi = x + a_1 r; \;\; \eta = y + a_2 r; \;\; \zeta = z + a_3 r,$$

and noting that $a_1^2 + a_2^2 + a_3^2 = 1$, we get:

$$r = \sqrt{(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2},$$

and the expression for u finally takes the form:

$$u(x, y, z, t) = \frac{1}{4\pi a^2} \iiint\limits_{r \leqslant at} \frac{f\left(\xi, \eta, \zeta, t - \dfrac{r}{a}\right)}{r} \, dv, \tag{91}$$

where the inequality $r \leqslant at$ characterizes the sphere D_{at} mentioned above. The characteristic feature of the integrand in the last expression is that the function f runs from the instant $t - (r/a)$ which precedes the instant t for which u is evaluated. The difference r/a between the instants gives the time required for passage from the point (ξ, η, ζ) to the point (x, y, z) with velocity a. Expression (91) is generally known as the *retarded potential*. It should also be noted that the basic formula (89) has a simple physical meaning: it shows that the solution of non-homogeneous equation (83) with initial conditions (84) is the sum of the impulses $w(x, y, z, t; \tau)d\tau$ derived from the presence of the term $f(x, y, z, t)$ and defined by equations (85) and (86).

We now consider the non-homogeneous wave equation for cylindrical waves:

$$\frac{\partial^2 u}{\partial t^2} = a^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + f(x, y, t) \tag{92}$$

with zero initial conditions. Precisely the same method as above can be used to obtain the solution in the form:

$$u(x, y, t) = \int\limits_0^t w(x, y, t; \tau) \, d\tau,$$

where $w(x, y, z, t; \tau)$ satisfies the homogeneous equation

$$\frac{\partial^2 w}{\partial t^2} = a^2\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right)$$

and the initial conditions

$$w\Big|_{t=\tau} = 0; \ \frac{\partial w}{\partial t}\Big|_{t=\tau} = f(x, y, \tau).$$

We finally get, on taking (80) into account:

$$u(x, y, t) = \frac{1}{2\pi a}\int_0^t \left[\iint\limits_{\varrho \leqslant a(t-\tau)} \frac{f(\xi, \eta, \tau)}{\sqrt{a^2(t-\tau)^2 - \varrho^2}}\, d\xi\, d\eta\right] d\tau \tag{93}$$

$$\left[\varrho^2 = (\xi - x)^2 + (\eta - y)^2\right].$$

It may be noted that we have integration with respect to time in this last formula, which was not the case with (91), where dependence on time reduced simply to the dependence on time of the radius of the sphere over which integration was carried out and the dependence on time of the function $f(x, y, z, t)$. In the linear case:

$$\frac{\partial^2 u}{\partial t^2} = a^2\frac{\partial^2 u}{\partial x^2} + f(x, t), \tag{94}$$

the solution obviously becomes:

$$u(x, t) = \frac{1}{2a}\int_0^t \left[\int_{x-a(t-\tau)}^{x+a(t-\tau)} f(\xi, \tau)\, d\xi\right] d\tau. \tag{95}$$

175. Point sources. We may suppose that the term $f(x, y, z, t)$ in equation (83) differs from zero only in a small sphere with centre at the coordinate origin. On letting the radius of the sphere tend to zero and the intensity of the external force increase indefinitely, we obtain in the limit the (so-called generalized) solution of the wave equation with the presence of a point source, which starts to act at the instant $t = 0$ and which can have any law of action in relationship to time. We take

$$f(x, y, z, t) = 0 \quad \text{for} \quad \sqrt{x^2 + y^2 + z^2} \geqslant \varepsilon \tag{96}$$

and

$$\iiint\limits_{C_\varepsilon} f(x, y, z, t)\, dx\, dy\, dz = 4\pi\omega(t), \tag{97}$$

where C_ε is a sphere with centre at the origin and radius ε. We return to equation (91), where we shall assume $at > \sqrt{x^2 + y^2 + z^2}$. By (96), it is sufficient to carry out the integration over the sphere C_ε. As

$\varepsilon \to 0$, r becomes equal in the limit to the distance from point (x, y, z) to the origin, i.e. $r = \sqrt{x^2 + y^2 + z^2}$, and we obtain on taking (97) into account:

$$u(x, y, z, t) = \frac{1}{r}\,\omega\left(t - \frac{r}{a}\right) \tag{98}$$

$$(r = \sqrt{x^2 + y^2 + z^2}).$$

Moreover, we must take $u(x, y, z, t) = 0$ for $r > at$, since the domain of integration in (91) for $r > at$ does not include the sphere C for sufficiently small ε. It may be noted that function (98) satisfies the homogeneous wave equation and has a singularity at the coordinate origin for any choice of twice continuously differentiable function $\omega(t)$.

In the case of equation (92), we proceed exactly as above and take

$$f(x, y, t) = 0 \quad \text{for} \quad \sqrt{x^2 + y^2} > \varepsilon,$$

and

$$\iint_{\gamma_\varepsilon} f(x, y, t)\,\mathrm{d}x\,\mathrm{d}y = 2\pi\omega(t),$$

where γ_ε is a circle of radius ε with centre at the origin. On returning to equation (93) and passing to the limit, we get the solution for a point source in the case of cylindrical waves as:

$$u(x, y, t) = \frac{1}{a}\int_0^{t - \frac{\varrho}{a}} \frac{\omega(\tau)}{\sqrt{a^2(t-\tau)^2 - \varrho^2}}\,\mathrm{d}\tau \qquad (at > \varrho), \tag{99}$$

$$u(x, y, t) = 0 \quad \text{for} \quad at < \varrho$$

$$(\varrho = \sqrt{x^2 + y^2}).$$

The difference between (98) and (99) is similar to the difference pointed out in the previous section. Formula (98) shows that the influence of a point source at the instant t at the point (x, y, z) depends only on the intensity of the source at the instant $t-(r/a)$. In the case of (99), this influence is determined by the action of the point source for the interval of time from $t = 0$ to $t = \varrho/a$.

In the linear case (94), after as usual putting

$$\int_{-\varepsilon}^{+\varepsilon} f(x, t)\,\mathrm{d}x = \omega(t) \quad \text{and} \quad f(x, t) = 0 \quad \text{for} \quad |x| > \varepsilon,$$

we obtain on passing to the limit in (95):

$$
\begin{aligned}
u(x,t) &= \int_0^{t-\frac{|x|}{a}} \omega(\tau)\,d\tau \quad \text{for} \quad |x| < at, \\
u(x,t) &= 0 \quad \text{for} \quad |x| > at.
\end{aligned} \tag{100}
$$

176. The transverse vibrations of a membrane. We have so far considered the wave equations in a plane and in space in the absence of a boundary, and have thus only had the initial conditions in addition to the differential equation. Boundary value problems involving the wave equation either in a plane or in space are much more difficult than in the linear case. We shall consider the two particular cases of such problems in a plane which arise when the boundary is a rectangle or a circle. The wave equation in a plane will be interpreted physically as the equation of the transverse vibrations of a membrane.

We understand by membrane a very thin sheet which, like a string, does work only by extension and not by bending. If the membrane is under the action of a uniform tension T_0 and lies in the (x, y) plane at equilibrium, and if we confine ourselves to the case when motion occurs parallel to axis OZ, the displacement u of a point (x, y) of the membrane becomes a function of x, y and t, satisfying a differential equation analogous to that for a string, i.e.

$$
\frac{\partial^2 u}{\partial t^2} = a^2\left(\frac{\partial^2 u}{\partial^2 x} + \frac{\partial^2 u}{\partial y^2}\right) + f(x, y, t), \tag{101}
$$

where

$$
a = \sqrt{\frac{T_0}{\varrho}},
$$

ϱ is the surface density of the membrane, and ϱf is the external force or loading. We shall not dwell here on the derivation of equation (101).

Apart from differential equation (101), we also have to take into account the *boundary conditions* which must be satisfied by the function u on the boundary of the membrane (C). We shall only discuss the case when the contour (C) is rigidly fixed, i.e.

$$
u = 0 \quad \text{on} \quad (C). \tag{102}
$$

Finally, we have to specify the initial conditions, i.e. the displacement and velocity of every point of the membrane at the initial instant:

$$
u\Big|_{t=0} = \varphi_1(x, y), \qquad \frac{\partial u}{\partial t}\Big|_{t=0} = \varphi_2(x, y). \tag{103}
$$

177. Rectangular membrane. We consider *the free vibrations of a rectangular membrane*, the contour of which consists of the rectangle with sides

$$
x = 0, \; x = l, \; y = 0, \; y = m \tag{104}
$$

in the (x, y) plane. We shall assume that external forces are absent, i.e. $f = 0$.

Here we have to find the solution of the equation

$$\frac{\partial^2 u}{\partial t^2} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \tag{105}$$

satisfying conditions (102) and (103).

On again applying the standing wave (Fourier's) method, we seek a particular solution of equation (105) in the form:

$$(a \cos \omega t + \beta \sin \omega t)\, U(x, y), \tag{106}$$

which gives us

$$-\omega^2 (a \cos \omega t + \beta \sin \omega t)\, U(x, y) = a^2 \left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2}\right)(a \cos \omega t + \beta \sin \omega t),$$

whence, on setting

$$\frac{\omega^2}{a^2} = k^2, \tag{107}$$

we obtain the equation for U as:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + k^2 U = 0.$$

We seek in turn a particular solution of this last equation in the form

$$U(x, y) = X(x)\, Y(y), \tag{108}$$

which gives us:

$$X''(x)\, Y(y) + X(x)\, Y''(y) + k^2 X(x)\, Y(y) = 0,$$

or

$$\frac{X''(x)}{X(x)} = -\frac{Y''(y) + k^2 Y(y)}{Y(y)} = -\lambda^2,$$

where λ^2 is an as yet undefined constant.

We thus have the two equations:

$$X''(x) + \lambda^2 X(x) = 0; \quad Y''(y) + \mu^2 Y(y) = 0, \tag{109}$$

where

$$\mu^2 = k^2 - \lambda^2, \quad \mu^2 + \lambda^2 = k^2.$$

Equations (109) give us the general forms of the functions $X(x)$ and $Y(y)$:

$$X(x) = C_1 \sin \lambda x + C_2 \cos \lambda x; \quad Y(y) = C_3 \sin \mu x + C_4 \cos \mu x.$$

The condition

$$u = 0 \quad \text{on} \quad (C)$$

gives us

$$U(x, y) = 0 \quad \text{on} \quad (C),$$

whilst this last condition in turn splits up into the following conditions:

$$X(0) = 0; \quad X(l) = 0; \quad Y(0) = 0; \quad Y(m) = 0,$$

from which it is obvious that $C_2 = C_4 = 0$; also, if we neglect the non-zero constant factors C_1 and C_3, we now have:

$$X(x) = \sin \lambda x, \quad Y(y) = \sin \mu y, \tag{110}$$

with, however,

$$\sin \lambda l = 0, \qquad \sin \mu m = 0. \tag{111}$$

It follows from equations (111) that λ and μ have the infinite set of values

$$\lambda = \lambda_1, \lambda_2, \ldots \lambda_\sigma, \ldots; \; \mu = \mu_1, \mu_2, \ldots \mu_\tau, \ldots; \; \lambda_\sigma = \frac{\sigma\pi}{l}, \; \mu_\tau = \frac{\tau\pi}{m}. \tag{112}$$

On arbitrarily selecting a λ and a μ from sequence (112), we get for the corresponding value of the constant k^2:

$$k^2_{\sigma,\tau} = \lambda^2_\sigma + \mu^2_\tau = \pi^2\left(\frac{\sigma^2}{l^2} + \frac{\tau^2}{m^2}\right),$$

and find from (107) the corresponding frequency ω for this value of k^2:

$$\omega^2_{\sigma,\tau} = a^2 k^2_{\sigma,\tau} = a^2\pi^2\left(\frac{\sigma^2}{l^2} + \frac{\tau^2}{m^2}\right). \tag{113}$$

We substitute λ_σ for λ and μ_τ for μ in expression (106) and write the corresponding α and β as $\alpha_{\sigma,\tau}$, $\beta_{\sigma,\tau}$ and hence obtain an infinite set of solutions of equation (105) satisfying boundary condition (102) in the form:

$$(\alpha_{\sigma,\tau} \cos \omega_{\sigma,\tau} t + \beta_{\sigma,\tau} \sin \omega_{\sigma,\tau} t) \sin \frac{\sigma\pi x}{l} \sin \frac{\tau\pi y}{m},$$

i.e. the infinite set of proper (free) harmonic vibrations of a membrane corresponding to the free vibrations of a string.

The constants α, β are determined from the initial conditions. Substitution of $t = 0$ in the expressions:

$$u = \sum_{\sigma,\tau=1}^{\infty} (\alpha_{\sigma,\tau} \cos \omega_{\sigma,\tau} t + \beta_{\sigma,\tau} \sin \omega_{\sigma,\tau} t) \sin \frac{\sigma\pi x}{l} \sin \frac{\tau\pi y}{m},$$

$$\frac{\partial u}{\partial t} = \sum_{\sigma,\tau=1}^{\infty} \omega_{\sigma,\tau} (\beta_{\sigma,\tau} \cos \omega_{\sigma,\tau} t - \alpha_{\sigma,\tau} \sin \omega_{\sigma,\tau} t) \sin \frac{\sigma\pi x}{l} \sin \frac{\tau\pi y}{m},$$

gives us on the basis of (103):

$$u\big|_{t=0} = \varphi_1(x, y) = \sum_{\sigma,\tau=1}^{\infty} \alpha_{\sigma,\tau} \sin \frac{\sigma\pi x}{l} \sin \frac{\tau\pi y}{m},$$

$$\frac{\partial u}{\partial t}\bigg|_{t=0} = \varphi_2(x, y) = \sum_{\sigma,\tau=1}^{\infty} \beta_{\sigma,\tau}\, \omega_{\sigma,\tau} \sin \frac{\sigma\pi x}{l} \sin \frac{\tau\pi y}{m}.$$

These expressions are in fact the double Fourier expansions of the functions φ_1 and φ_2, and it may readily be seen that the coefficients α and β are given by

$$\left.\begin{aligned} \alpha_{\sigma,\tau} &= \frac{4}{lm}\int_0^l\int_0^m \varphi_1(\xi,\zeta)\sin\frac{\sigma\pi\xi}{l}\sin\frac{\tau\pi\eta}{m}\,\mathrm{d}\xi\,\mathrm{d}\eta\,, \\ \omega_{\sigma,\tau}\beta_{\sigma,\tau} &= \frac{4}{lm}\int_0^l\int_0^m \varphi_2(\xi,\eta)\sin\frac{\sigma\pi\xi}{l}\sin\frac{\tau\pi\eta}{m}\,\mathrm{d}\xi\,\mathrm{d}\eta\,, \end{aligned}\right\} \tag{114}$$

which provides the solution of our problem.

The case of a membrane differs from that of a string inasmuch as, for the latter, each proper frequency corresponds to a particular form of the string, which is simply split up by nodes into several different parts. It is possible with a membrane that the same frequency corresponds to several different shapes with different positions of the *nodal lines*, i.e. the lines on which the amplitude of vibration becomes zero. This can be studied most simply by considering a square membrane:

$$l = m = r.$$

The frequency $\omega_{\sigma,\tau}$ is defined in this case by

$$\omega_{\sigma,\tau} = \frac{a\pi}{r}\sqrt{\sigma^2+\tau^2} = \alpha\sqrt{\sigma^2+\tau^2}, \tag{115}$$

where $\alpha = a\pi/r$ is a factor depending neither on σ nor τ.

On setting $\sigma = \tau = 1$, we get the fundamental tone u_{11} of the membrane with frequency $\omega_{11} = \alpha\sqrt{2}$:

$$u_{11} = N_1 \sin(\omega_{11}t + \varphi_{11})\sin\frac{\pi x}{r}\sin\frac{\pi y}{r}.$$

Here, there are no nodal lines whatever within the membrane.

On next putting

$$\sigma = 1,\quad \tau = 2 \quad\text{or}\quad \sigma = 2,\quad \tau = 1,$$

we get two further tones of the same frequency

$$\omega_{12} = \omega_{21} = \alpha\sqrt{5}\,,$$

given by:

$$u_{12} = N_{12}\sin(\omega_{12}t + \varphi_{12})\sin\frac{\pi x}{r}\sin\frac{2\pi y}{r}\,,$$

$$u_{21} = N_{21}\sin(\omega_{21}t + \varphi_{21})\sin\frac{2\pi x}{r}\sin\frac{\pi y}{r}.$$

The nodal lines of these elementary vibrations are respectively

$$y = \frac{r}{2} \quad\text{or}\quad x = \frac{r}{2}.$$

But in addition to vibrations u_{12} and u_{21}, an infinite set of further vibrations of the same frequency exists, obtained from linear combinations of u_{12} and u_{21}. On writing $\varphi_{12} = \varphi_{21} = 0$ for simplicity, we get a vibration of the form

$$\sin \omega t \left[N_1 \sin \frac{\pi x}{r} \sin \frac{2\pi y}{r} + N_2 \sin \frac{2\pi x}{r} \sin \frac{\pi y}{r} \right],$$

where $\omega - \omega_{12}\, \omega_{21}$, $N_1 = N_{12}$ and $N_2 = N_{21}$.

With $N_1 = N_2$, the nodal lines are defined by the equation

$$0 = \sin \frac{\pi x}{r} \sin \frac{2\pi y}{r} + \sin \frac{2\pi x}{r} \sin \frac{\pi y}{r} = 2 \sin \frac{\pi x}{r} \sin \frac{\pi y}{r} \left(\cos \frac{\pi x}{r} + \cos \frac{\pi y}{r} \right),$$

giving us the nodal line

$$x + y = r.$$

With $N_2 = -N_1$, we find by precisely the same method the nodal line $x - y = 0$.

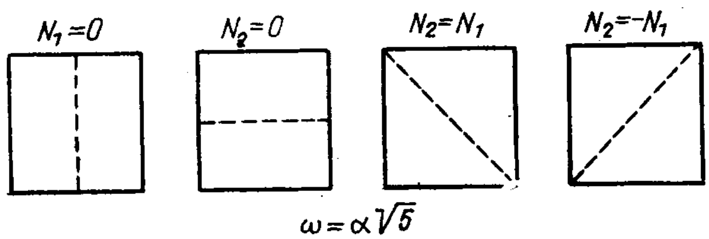

FIG. 134

These elementary cases are illustrated in Fig. 134. We get more complicated nodal lines for the same frequency when $N_2 \neq \pm N_1$ and $N_1, N_2 \neq 0$.

These all have equations of the form

$$N_2 \cos \frac{\pi x}{r} + N_1 \cos \frac{\pi y}{r} = 0.$$

On now setting

$$\sigma = 2, \quad \tau = 2,$$

we obtain the unique tone of frequency

$$\omega_{22} = a\sqrt{8},$$

the nodal lines of which are (Fig. 135):

$$x = \frac{1}{2} r \quad \text{and} \quad y = \frac{1}{2} r.$$

The next case:

$$\sigma = 1, \quad \tau = 3, \quad \sigma = 3, \quad \tau = 1$$

again leads to an infinite set of vibrations of the same frequency $\omega_{13} = \omega_{31} = a\sqrt{10}$. Figure 136 illustrates their nodal lines in elementary cases analogous to those with frequency $\omega_{12} = \omega_{21} = a\sqrt{5}$. All these figures are in fact the well known Chladni's figures of acoustics.

The *forced vibrations of a membrane* are investigated in exactly the same way as the forced vibrations of a string, except that the external force $f(x, y, t)$ is expanded in a double instead of a simple Fourier series.

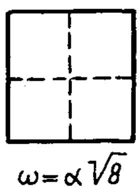

FIG. 135

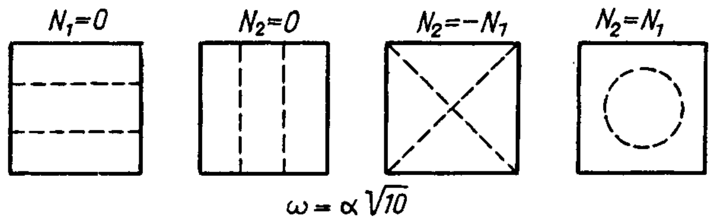

FIG. 136

178. Circular membranes. The case of a circular membrane gives us an example of the expansion of a given function in Bessel functions; it is important inasmuch as similar expansions are encountered in other notable problems of mathematical physics.

We investigate the free (proper) vibrations of a circular membrane with contour of radius l and centre at the coordinate origin. We assume as a preliminary that there is no displacement on the contour. On introducing polar coordinates (r, θ) instead of rectangular coordinates (x, y), we have now

$$u\,|_{r=l} = 0.$$

As in the rectangular case, we seek particular solutions of equation (105) in the form:

$$(a \cos \omega t + \beta \sin \omega t)\, U,$$

where the function U is now assumed given in terms of (r, θ) instead of (x, y) We obtain the same differential equation for U:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + k^2 U = 0, \tag{116}$$

except that it has to be transformed to the new variables (r, θ). This is done simply by expressing Laplace's operator

$$\Delta U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} \tag{117}$$

in polar coordinates. We know that the operator for three variables

$$\Delta U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}$$

is given in terms of cylindrical coordinates by [119]:

$$\Delta U = \frac{1}{\varrho}\left[\frac{\partial}{\partial \varrho}\left(\varrho \frac{\partial U}{\partial \varrho}\right) + \frac{1}{\varrho}\frac{\partial^2 U}{\partial \varphi^2} + \varrho \frac{\partial^2 U}{\partial z^2}\right]$$

We express (117) in polar coordinates by taking U as independent of z. We shall in future denote the length of the radius vector as r instead of ϱ, and the polar angle by θ instead of φ:

$$\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}=\frac{\partial^2 U}{\partial r^2}+\frac{1}{r}\frac{\partial U}{\partial r}+\frac{1}{r^2}\frac{\partial^2 U}{\partial \theta^2}.$$

Equation (116) becomes:

$$\frac{\partial^2 U}{\partial r^2}+\frac{1}{r}\frac{\partial U}{\partial r}+\frac{1}{r^2}\frac{\partial^2 U}{\partial \theta^2}+k^2U=0.$$

We seek its particular solution as the product:

$$U(r,\theta)=T(\theta)\cdot R(r)$$

which gives:

$$T(\theta)\left[R''(r)+\frac{1}{r}R'(r)+k^2R(r)\right]+\frac{1}{r^2}T''(\theta)\,R(r)=0,$$

or

$$\frac{T''(\theta)}{T(\theta)}=-\frac{r^2R''(r)+rR'(r)+k^2r^2R(r)}{R(r)}=-\lambda^2$$

and so finally:

$$T''(\theta)+\lambda^2\,T(\theta)=0, \tag{118}$$

$$R''(r)+\frac{1}{r}R'(r)+\left(k^2-\frac{\lambda^2}{r^2}\right)R(r)=0. \tag{119}$$

The general solution of equation (118) is of the form

$$T(\theta)=C\cos\lambda\theta+D\sin\lambda\theta,$$

and since the sense of the actual problem implies that U must be a single-valued periodic function of θ of period 2π, we can say the same of the function $T(\theta)$, which in turn implies that λ must be an integer. If we confine ourselves to positive λ, we have to take $\lambda=0,1,2,\ldots,n,\ldots,$ the corresponding expressions for $T(\theta)$ and $R(r)$ being denoted by

$$T_0(\theta),\ T_1(\theta),\ T_2(\theta),\ \ldots,\ T_n(\theta),\ \ldots;\ R_0(r),\ R_1(r),\ R_2(r),\ \ldots,\ R_n(r),\ \ldots$$

We thus obtain an infinity of solutions of (105) of the form

$$(\alpha\cos\omega t+\beta\sin\omega t)(C\cos n\theta+D\sin n\theta)\,R_n(r)\quad(\omega=ak). \tag{120}$$

The function $R_n(r)$ satisfies equation (119) if we replace the λ there by n:

$$R_n''(r)+\frac{1}{r}R_n'(r)+\left(k^2-\frac{n^2}{r^2}\right)R_n(r)=0. \tag{121}$$

As we saw in [49], the general solution of this equation is

$$R_n(r)=C_1J_n(kr)+C_2K_n(kr), \tag{122}$$

where $J_n(x)$ is the Bessel function and $K_n(x)$ is the second solution of Bessel's equation which tends to infinity at $x=0$; since the nature of the problem implies that the required solutions remain bounded at all points of the membrane,

the origin $r = 0$ included, the term in $K_n(kr)$ must be absent in the above expression for $R_n(r)$, i.e. $C_2 = 0$. We can take $C_1 = 1$ without loss of generality, i.e. we put

$$R_n(r) = J_n(kr), \tag{123}$$

in which case the boundary condition

$$u|_{r=l} = 0$$

gives

$$J_n(kl) = 0. \tag{124}$$

On writing $kl = \mu$, we have the transcendental equation for μ:

$$J_n(\mu) = 0, \tag{125}$$

which, as shown in the theory of Bessel functions, has the infinite set of positive roots

$$\mu_1^{(n)},\ \mu_2^{(n)},\ \mu_3^{(n)},\ \ldots,\ \mu_m^{(n)},\ \ldots, \tag{126}$$

corresponding to the values

$$k_1^{(n)},\ k_2^{(n)},\ k_3^{(n)},\ \ldots;\ k_m^{(n)} = \frac{\mu_m^{(n)}}{l} \tag{127}$$

of the parameter k and, by (107), to the values

$$\omega_{m,n} = ak_m^{(n)} \quad (n = 0,\ 1,\ 2,\ \ldots,\ m = 1.\ 2,\ \ldots) \tag{128}$$

of the frequency ω. The first nine roots of the first six Bessel functions are given in the attached table:

1	2.404	3.832	5.135	6.379	7.586	8.780
2	5.520	7.016	8.417	9.760	11.064	12.339
3	8.654	10.173	11.620	13.017	14.373	15.700
4	11.792	13.323	14.796	16.224	17.616	18.982
5	14.931	16.470	17.960	19.410	20.827	22.220
6	18.076	19.616	21.117	22.583	24.018	25.431
7	21.212	22.760	24.270	25.749	27.200	28.628
8	24.353	25.903	27.421	28.909	30.371	31.813
9	27.494	29.047	30.571	32.050	33.512	34.983

Subsequent roots may be evaluated from the approximate formula

$$k_m^{(n)} = \frac{1}{4}\pi(2n - 1 + 4m) - \frac{4n^2 - 1}{\pi(2n - 1 + 4m)}, \tag{129}$$

the accuracy of which, for a given n, increases with m. We cannot enter here into the derivation of formula (129).

It follows from expression (120) that the particular solutions obtained by us can be written in the form:

$$\left(\alpha^{(1)}_{m,n}\cos\omega_{m,n}t+\alpha^{(2)}_{m,n}\sin\omega_{m,n}t\right)\cos n\theta\cdot J_n\left(k^{(n)}_m r\right)+$$
$$+\left(\beta^{(1)}_{m,n}\cos\omega_{m,n}t+\beta^{(2)}_{m,n}\sin\omega_{m,n}t\right)\sin n\theta\cdot J_n\left(k^{(n)}_m r\right) \qquad (130)$$
$$(m,\ n=1,\ 2,\ \ldots).$$

We also notice that equation (118) with $\lambda=0$ has two solutions: a constant and θ. The second solution is unsuitable since it is not periodic. Expression (120) gives in the first case the solution:

$$\left(\alpha^{(1)}_{m,0}\cos\omega_{m,0}t+\alpha^{(2)}_{m,0}\sin\omega_{m,0}t\right)J_0\left(k^{(0)}_m r\right).$$

This solution also has the form (130) (with $n=0$), the only difference being that with $n=0$ the second term in (130) vanishes due to the presence of the factor $\sin n\theta$.

It only remains for us now to satisfy the initial conditions:

$$u\,|_{t=0}=\varphi_1(r,\theta)\,;\quad \left.\frac{\partial u}{\partial t}\right|_{t=0}=\varphi_2(r,\theta). \qquad (131)$$

To this end, on bearing in mind the particular solutions arrived at, we seek u as the double series:

$$u(r,\theta,t)=\sum_{\substack{n=0\\m=1}}^{\infty}\left(\alpha^{(1)}_{m,n}\cos\omega_{m,n}t+\alpha^{(2)}_{m,n}\sin\omega_{m,n}t\right)\cos n\theta\cdot J_n\left(k^{(n)}_m r\right)+$$
$$+\sum_{\substack{n=0\\m=1}}^{\infty}\left(\beta^{(1)}_{m,n}\cos\omega_{m,n}t+\beta^{(2)}_{m,n}\sin\omega_{m,n}t\right)\sin n\theta\cdot J_n\left(k^{(n)}_m r\right).$$

On evaluating $\partial u/\partial t$:

$$\frac{\partial u}{\partial t}=\sum_{\substack{n=0\\m=1}}^{\infty}\omega_{m,n}\left(\alpha^{(2)}_{m,n}\cos\omega_{m,n}t-\alpha^{(1)}_{m,n}\sin\omega_{m,n}t\right)\cos n\theta\cdot J_n\left(k^{(n)}_m r\right)+$$
$$+\sum_{\substack{n=0\\m=1}}^{\infty}\omega_{m,n}\left(\beta^{(2)}_{m,n}\cos\omega_{m,n}t-\beta^{(1)}_{m,n}\sin\omega_{m,n}t\right)\sin n\theta\cdot J_n\left(k^{(n)}_m r\right),$$

and setting $t=0$ in these expressions, we arrive by (131) at the point of having to expand the given functions $\varphi_1(r,\theta)$ and $\varphi_2(r,\theta)$ as double series of the forms:

$$\left.\begin{aligned}\varphi_1(r,\theta)&=\sum_{\substack{n=0\\m=1}}^{\infty}\left(\alpha^{(1)}_{m,n}\cos n\theta+\beta^{(1)}_{m,n}\sin n\theta\right)\cdot J_n\left(k^{(n)}_m r\right),\\ \varphi_2(r,\theta)&=\sum_{\substack{n=0\\m=1}}^{\infty}\omega_{m,n}\left(\alpha^{(2)}_{m,n}\cos n\theta+\beta^{(2)}_{m,n}\sin n\theta\right)\cdot J_n\left(k^{(n)}_m r\right).\end{aligned}\right\} \qquad (132)$$

Expansion of $\varphi_1(r, \theta)$ as a periodic function of θ in an ordinary Fourier series gives us

$$\varphi_1(r, \theta) = \frac{\varphi_0^{(1)}}{2} + \sum_{n=1}^{\infty} \left(\varphi_n^{(1)} \cos n\theta + \psi_n^{(1)} \sin n\theta\right),$$

where

$$\varphi_n^{(1)} = \frac{1}{\pi} \int_{-\pi}^{\pi} \varphi_1(r, \theta) \cos n\theta \, d\theta \; ; \; \psi_n^{(1)} = \frac{1}{\pi} \int_{-\pi}^{\pi} \varphi_1(r, \theta) \sin n\theta \, d\theta \quad (n=0, 1, 2, \ldots). \tag{133}$$

On comparing this expansion with the first of expressions (132), we readily obtain:

$$\left.\begin{gathered} \varphi_0^{(1)} = 2 \sum_{m=1}^{\infty} \alpha_{m,0}^{(1)} J_0\left(k_m^{(0)} r\right); \quad \varphi_n^{(1)} = \sum_{m=1}^{\infty} \alpha_{m,n}^{(1)} J_n\left(k_m^{(n)} r\right); \\ \psi_n^{(1)} = \sum_{m=1}^{\infty} \beta_{m,n}^{(1)} J_n\left(k_m^{(n)} r\right). \end{gathered}\right\} \tag{134}$$

Coefficients $\varphi^{(1)}$ and $\psi^{(1)}$ obviously depend on r, as indicated by their expressions (133). We thus arrive at the problem of expanding a given function of r in a series in $J_n(k_m^{(n)}r)$ with fixed n. On obtaining these expansions, we can determine the coefficients α and β, and the problem is solved.

Thus, let it be required to expand a given function $f(r)$ as

$$f(r) = \sum_{m=1}^{\infty} A_m J_n\left(k_m^{(n)} r\right). \tag{135}$$

We assume that this expansion is possible and that it can be integrated term by term, so that we only have to show how to find the coefficients A_m. To this end, we prove that the functions

$$J_n\left(k_1^{(n)} r\right), \quad J_n\left(k_2^{(n)} r\right), \; \ldots, \; J_n\left(k_m^{(n)} r\right), \; \ldots$$

possess the property of *generalized orthogonality*, i.e.

$$\int_0^l J_n\left(k_\sigma^{(n)} r\right) J_n\left(k_\tau^{(n)} r\right) r \, dr = 0 \quad \text{for} \quad \sigma \neq \tau. \tag{136}$$

In fact, if we substitute $k_\sigma^{(n)2}$ then $k_\tau^{(n)2}$ for k^2, and similarly $J_n(k_\sigma^{(n)} r)$ then $J_n(k_\tau^{(n)} r)$ for $R_n(r)$ in equation (121), we get:

$$\frac{d^2 J_n\left(k_\sigma^{(n)} r\right)}{dr^2} + \frac{1}{r} \frac{dJ_n\left(k_\sigma^{(n)} r\right)}{dr} + \left(k_\sigma^{(n)2} - \frac{n^2}{r^2}\right) J_n\left(k_\sigma^{(n)} r\right) = 0,$$

$$\frac{d^2 J_n\left(k_\tau^{(n)} r\right)}{dr^2} + \frac{1}{r} \frac{dJ_n\left(k_\tau^{(n)} r\right)}{dr} + \left(k_\tau^{(n)2} - \frac{n^2}{r^2}\right) J_n\left(k_\tau^{(n)} r\right) = 0.$$

On multiplying the first equation by $rJ_n(k_\tau^{(n)} r)$ and the second equation by $rJ_n(k_\sigma^{(n)} r)$, subtracting, then integrating from 0 to l, we get:

$$\left(k_\sigma^{(n)2} - k_\tau^{(n)2}\right) \int_0^l J_n\left(k_\sigma^{(n)} r\right) J_n\left(k_\tau^{(n)} r\right) r \, \mathrm{d}r =$$

$$= \int_0^l \left[\frac{\mathrm{d}^2 J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r^2} J_n\left(k_\sigma^{(n)} r\right) - \frac{\mathrm{d}^2 J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r^2} J_n\left(k_\tau^{(n)} r\right)\right] r \, \mathrm{d}r +$$

$$+ \int_0^l \left[\frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} J_n\left(k_\sigma^{(n)} r\right) - \frac{\mathrm{d}J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r} J_n\left(k_\tau^{(n)} r\right)\right] \mathrm{d}r.$$

We have by integrating by parts:

$$\int \frac{\mathrm{d}^2 J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r^2} J_n\left(k_\sigma^{(n)} r\right) r \, \mathrm{d}r = \frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} rJ_n\left(k_\sigma^{(n)} r\right) -$$

$$- \int \frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} \cdot \frac{\mathrm{d}\left[rJ_n\left(k_\sigma^{(n)} r\right)\right]}{\mathrm{d}r} \mathrm{d}r = \frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} rJ_n\left(k_\sigma^{(n)} r\right) -$$

$$- \int \frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} \cdot \frac{\mathrm{d}J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r} r \, \mathrm{d}r - \int \frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} J_n\left(k_\sigma^{(n)} r\right) \mathrm{d}r,$$

and similarly:

$$\int \frac{\mathrm{d}^2 J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r^2} J_n\left(k_\tau^{(n)} r\right) r \, \mathrm{d}r = \frac{\mathrm{d}J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r} rJ_n\left(k_\tau^{(n)} r\right) -$$

$$- \int \frac{\mathrm{d}J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r} \cdot \frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} r \, \mathrm{d}r - \int \frac{\mathrm{d}J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r} J_n\left(k_\tau^{(n)} r\right) \mathrm{d}r.$$

We easily deduce from the above that

$$\left(k_\sigma^{(n)2} - k_\tau^{(n)2}\right) \int_0^l J_n\left(k_\sigma^{(n)} r\right) J_n\left(k_\tau^{(n)} r\right) r \mathrm{d}r =$$

$$= r \left[\frac{\mathrm{d}J_n\left(k_\tau^{(n)} r\right)}{\mathrm{d}r} J_n\left(k_\sigma^{(n)} r\right) - \frac{\mathrm{d}J_n\left(k_\sigma^{(n)} r\right)}{\mathrm{d}r} J_n\left(k_\tau^{(n)} r\right)\right] \Bigg|_{r=0}^{r=l}.$$

We have by definition of $k_\sigma^{(n)}$, $k_\tau^{(n)}$:

$$J_n\left(k_\sigma^{(n)} l\right) = J_n\left(k_\tau^{(n)} l\right) = 0,$$

whence it follows that the right-hand side of the equation written vanishes for $r = l$. In view of the presence of the factor r and the finiteness of $J_n(x)$ and $J_n'(x)$ for $x = 0$, we can say that the right-hand side also vanishes at the lower limit $r = 0$; but now it follows from the fact that $k_\sigma^{(n)} \neq k_\tau^{(n)}$ for $\sigma \neq \tau$ that

$$\int_0^l J_n\left(k_\sigma^{(n)} r\right) J_n\left(k_\tau^{(n)} r\right) r \, \mathrm{d}r = 0,$$

which is what we wished to prove.

Having proved (136), the determination of the coefficients A_m in expansion (135) presents no difficulty: we multiply both sides of equation (135) by $J_n(k_p^{(n)} r)$, integrate from 0 to l with respect to r, use formula (136), and find directly that:

$$\int_0^l f(r)\, J_n\left(k_p^{(n)} r\right) r\, \mathrm{d}r = A_p \int_0^l J_n^2\left(k_p^{(n)} r\right) r\, \mathrm{d}r.$$

Hence we can say that, if expansion (135) exists and is integrable term by term, the coefficients A_m are given by:

$$A_m = \frac{\int_0^l f(r)\, J_n\left(k_m^{(n)} r\right) r\, \mathrm{d}r}{\int_0^l J_n^2\left(k_m^{(n)} r\right) r\, \mathrm{d}r}.$$

Expressions (133) and (134) now give us the following formulae for the coefficients $\alpha^{(1)}$ and $\beta^{(1)}$:

$$\alpha_{m,0}^{(1)} = \frac{1}{2}\, \frac{\int_0^l \varphi_0^{(1)} J_0\left(k_m^{(0)} r\right) r\, \mathrm{d}r}{\int_0^l J_n^2\left(k_m^{(0)} r\right) r\, \mathrm{d}r} = \frac{1}{2\pi \int_0^l J_0^2\left(k_m^{(0)} r\right) r\, \mathrm{d}r} \int_{-\pi}^{\pi} \mathrm{d}\theta \int_0^l \varphi_1(r, \theta)\, J_0\left(k_m^{(0)} r\right) r\, \mathrm{d}r$$

$$\alpha_{m,n}^{(1)} = \frac{1}{\pi \int_0^l J_n^2\left(k_m^{(n)} r\right) r\, \mathrm{d}r} \int_{-\pi}^{\pi} \mathrm{d}\theta \int_0^l \varphi_1(r, \theta) \cos n\theta J_n\left(k_m^{(n)} r\right) r\, \mathrm{d}r$$

$$\beta_{m,n}^{(1)} = \frac{1}{\pi \int_0^l J_n^2\left(k_m^{(n)} r\right) r\, \mathrm{d}r} \int_{-\pi}^{\pi} \mathrm{d}\theta \int_0^l \varphi_1(r, \theta) \sin n\theta J_n\left(k_m^{(n)} r\right) r\, \mathrm{d}r.$$

The same arguments can be used for determining the coefficients $\alpha^{(2)}$, $\beta^{(2)}$; it is only necessary to replace φ_1 by φ_2 in the above formulae and divide the corresponding expressions by $\omega_{m,n}$.

As in the case of a rectangular membrane, the general motion of a circular membrane consists of the addition of an infinite set of proper harmonic vibrations, it being possible for the same frequency to correspond to an infinity of different dispositions of the nodal lines. Several dispositions of the nodal lines are illustrated in Fig. 137, along with the corresponding frequencies, the fundamental frequency being taken as unity; the radii of the circular nodal lines are also given, as fractions of the radius of the membrane.

When we apply Fourier's method for the case of any contour, we can only separate out the factor depending on t in (106), which leads to the equation:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + k^2 U = 0, \tag{137}$$

and we have to find the values of the parameter k for which the equation written has non-zero solutions which also satisfy the boundary condition (102). We succeeded in doing this in the above examples with the aid of further expansion of the variables. This method cannot be used in the general case, and we have to consider equation (137) directly. Of course the problem does not have an explicit solution. The theoretical solution and some relevant qualitative results are given in Volume IV. The boundary value problem for the wave equation in threedimensional space in the case of a rectangular parallelepiped is solved precisely as in [177], except that we arrive at Fourier series in three variables x, y, z. The case of a sphere again leads to Bessel functions. We shall discuss this in Volume III in connection with a more detailed treatment of the theory of Bessel functions.

A detailed study of the convergence of the Fourier series obtained in the solution of boundary value problems for the wave equation in the case of several spatial variables is given in Volume IV.

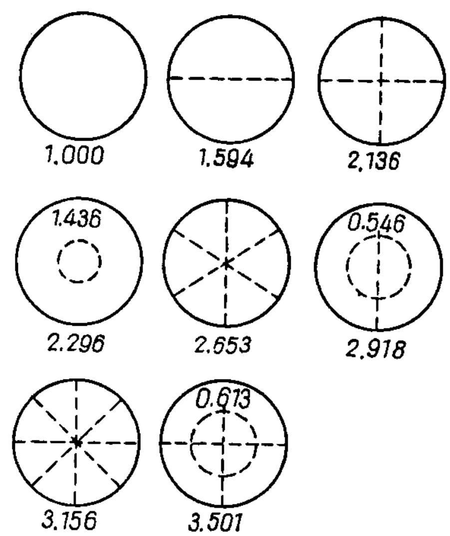

FIG. 137

179. The uniqueness theorem. We now prove the uniqueness of the solution of the wave equation both in the case of unbounded space with given initial conditions, and with the further imposition of boundary conditions. For simplicity of writing we shall take the velocity $a = 1$, which is permissible on replacing the t in the wave equation by at. We take the case of three independent variables for the sake of definiteness, i.e. the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, \tag{138}$$

and we start by considering the problem with initial conditions only, given throughout the (x, y) plane:

$$u\big|_{t=0} = \varphi(x, y); \quad \frac{\partial u}{\partial t}\bigg|_{t=0} = \varphi_1(x, y). \tag{139}$$

The solution of this problem has already been described [172], and the method used could itself serve for proving uniqueness. Here we shall give a different proof, which is also applicable to problems with boundary conditions. If equation (138) with initial conditions (139) has

two solutions, u_1 and u_2, their difference $(u_2 - u_1)$ must satisfy (138) and the homogeneous initial conditions:

$$u\Big|_{t=0} = 0; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = 0. \tag{140}$$

We now have to show that u must be identically zero for any (x, y) and for any $t > 0$. We consider a three-dimensional space (x, y, t) and take a point $N(x_0, y_0, t_0)$ of the space such that $t_0 > 0$. From this point as vertex, we draw the conical surface

$$(x - x_0)^2 + (y - y_0)^2 - (t - t_0)^2 = 0 \tag{141}$$

to its intersection with the plane $t = 0$. We also draw the plane $t = t_1$, where $0 < t_1 < t_0$, and let D be the three-dimensional domain bounded by the lateral surface Γ of our cone and the pieces of the planes $t = 0$ and $t = t_1$ lying inside the cone (D is a section of a cone). We can easily verify the following elementary identity:

$$2\frac{\partial u}{\partial t}\left(\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2}\right) = \frac{\partial}{\partial t}\left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial t}\right)^2\right] -$$
$$- 2\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t}\frac{\partial u}{\partial x}\right) - 2\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial t}\frac{\partial u}{\partial y}\right). \tag{142}$$

We integrate both sides over the domain D. The integral of the left-hand side must vanish since u is a solution of (138). We can use Ostrogradskii's formula to transform the integral of the right-hand side to an integral over the surface of D:

$$\iint\left\{\left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial t}\right)^2\right]\cos(n, t) -\right.$$
$$\left. - 2\frac{\partial u}{\partial t}\frac{\partial u}{\partial x}\cos(n, x) - 2\frac{\partial u}{\partial t}\frac{\partial u}{\partial y}\cos(n, y)\right\} ds. \tag{143}$$

The function u and all its first order partial derivatives are zero by (140) on the lower base of the section of cone D and integral (143) vanishes over the lower base. We have on the upper base σ:

$$\cos(n, x) = \cos(n, y) = 0 \text{ and } \cos(n, t) = 1,$$

On the lateral surface of the cone Γ the direction-cosines of the normals satisfy the relationship:

$$\cos^2(n, t) - \cos^2(n, x) - \cos^2(n, y) = 0,$$

and we can rewrite integral (143) over Γ in the form:

$$J = \iint\limits_{\Gamma} \frac{1}{\cos(n,t)} \left\{ \left[\frac{\partial u}{\partial x} \cos(n,t) - \frac{\partial u}{\partial t} \cos(n,x) \right]^2 + \right.$$

$$\left. + \left[\frac{\partial u}{\partial y} \cos(n,t) - \frac{\partial u}{\partial t} \cos(n,y) \right]^2 \right\} ds,$$

so that we finally get

$$J + \iint\limits_{\sigma_{t_1}} \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial u}{\partial t} \right)^2 \right] ds = 0.$$

We have $\cos(n, t) > 0$ and consequently $J \geqslant 0$ on the surface Γ, so that

$$\iint\limits_{\sigma_{t_1}} \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \frac{\partial u}{\partial t} \Big)^2 \right] ds = 0,$$

from which it follows that at all points inside the total cone with vertex $N(x_0, y_0, t_0)$ the first order partial derivatives of u vanish, with the result that u is constant. It is zero on the bases of the cone by (140), and it is therefore also zero at the point N. This proof of the uniqueness theorem can readily be extended to the case of a boundary problem with equation (138). Let the solution of (138) be sought in a domain B of the (x, y) plane with given initial and boundary value conditions, the boundary conditions being in reference to a contour l of domain B. We construct a cylinder with base B and generators parallel to the t axis. Every point of this cylinder corresponds to a definite point of domain B and a definite instant t. We suppose that we have the zero initial data (140) in B and that on the contour l of B we have the homogeneous boundary condition:

$$u|_l = 0. \tag{144}$$

We show that the function u is zero at all points of our cylinder. We take one such point N and draw the cone (141) through it. Let D be the solid bounded by the lateral surface of this cone, by our cylinder and by the planes $t = 0$ and $t = t_1$. We again integrate both sides of identity (142) over this domain. All the arguments remain the same except for the appearance of the integral over the lateral surface of the cylinder on the right-hand side. If this integral can be shown to vanish, the above proof of the uniqueness theorem can be retained complete. The integrand in this integral coincides with the integrand of (143). But we have $\cos(n, t) = 0$, and furthermore, $\partial u/\partial t = 0$, on the lateral surface of the cylinder. The last equation follows at once

from the fact that points of the lateral surface of the cylinder consist of points of the contour l at different instants t, whilst we have the homogeneous boundary condition (144) on the contour l for any t. The integrand of (143) thus vanishes at all points of the lateral surface of the cylinder, and the above proof of the uniqueness theorem is fully preserved for the boundary value problem. During the course of the proof, we integrated the right-hand side of expression (142) over the domain D and applied Ostrogradskii's formula. These operations are entirely justifiable if we assume that the function u has continuous derivatives up to the second order, which remain bounded inside the domain D.

We mentioned above that the investigation of problems of practical interest forces us to bring in so-called generalized solutions. We shall show in Volume IV that the uniqueness theorem is also valid for this wider class of generalized solutions.

180. Applications of Fourier integrals. We take the wave equation in the linear case:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, \tag{145}$$

for the semi-infinite domain $x > 0$, with the initial conditions

$$u\,|_{t=0} = \varphi(x) \qquad \left.\frac{\partial u}{\partial t}\right|_{t=0} = \varphi_1(x) \qquad (x > 0) \tag{146}$$

and the boundary condition

$$u\,|_{x=0} = 0. \tag{147}$$

This problem may easily be solved by the method indicated in [166]. In fact, it is sufficient to make an odd continuation of functions $\varphi(x)$ and $\varphi_1(x)$, assigned in the interval $(0, +\infty)$, into the interval $(-\infty, +\infty)$, and afterwards to apply formula (17) for an infinite string. We obtain by setting $x = 0$ in this formula:

$$u\,|_{x=0} = \frac{\varphi(-at) + \varphi(at)}{2} + \int_{-at}^{+at} \varphi_1(z)\,dz,$$

and both terms vanish in view of the odd continuation of $\varphi(z)$ and $\varphi_1(z)$, so that the boundary condition is in fact satisfied.

If we apply Fourier's method to the above problem, we get a Fourier integral instead of a Fourier series. As we saw in [167], application of Fourier's method with the boundary condition taken into account leads to a solution of the form:

$$u = (A \cos akt + B \sin akt) \sin kx.$$

There is no second boundary condition, so that all values are permissible for the parameter k, i.e. we arrive at a continuous spectrum of possible frequencies k of the semi-infinite string. Instead of summing over discrete values of k as in [167], we now have to integrate with respect to the parameter k, A and B being of course taken as functions of k. We obtain:

$$u(x, t) = \int_{-\infty}^{+\infty} [A(k) \cos akt + B(k) \sin akt] \sin kx \, dk. \tag{148}$$

The functions $A(k)$ and $B(k)$ must be found from the initial conditions (146) these give:

$$\varphi(x) = \int_{-\infty}^{+\infty} A(k) \sin kx \, dk; \quad \varphi_1(x) = \int_{-\infty}^{+\infty} ak\, B(k) \sin kx \, dk. \tag{149}$$

On comparing these formulae with Fourier's formula for an odd function:

$$f(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \left[\int_0^{\infty} f(t) \sin at \, dt \right] \sin ax \, da,$$

we obtain for $A(k)$ and $B(k)$:

$$A(k) = \frac{1}{\pi} \int_0^{\infty} \varphi(\xi) \sin k\xi \, d\xi; \qquad B(k) = \frac{1}{\pi ak} \int_0^{\infty} \varphi_1(\xi) \sin k\xi \, d\xi,$$

and substitution in (148) gives us the solution of the problem as

$$u(x, t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \left\{ \int_0^{\infty} \left[\varphi(\xi) \cos akt + \frac{1}{ak} \varphi_1(\xi) \sin akt \right] \sin k\xi \sin kx \, d\xi \right\} dk,$$

or alternatively, if we take into account the evenness of the integrand considered as a function of k, as:

$$u(x, t) = \frac{2}{\pi} \int_0^{\infty} \left\{ \int_0^{\infty} \left[\varphi(\xi) \cos akt + \frac{1}{ak} \varphi_1(\xi) \sin akt \right] \sin k\xi \, d\xi \right\} \sin kx \, dk.$$

It may readily be seen by using Fourier's formula that the right-hand side of the above expression coincides with the right-hand side of (17), on the hypothesis that $\varphi(x)$ and $\varphi_1(x)$ are odd.

In precisely the same way, we can consider a boundary value problem for the equation

$$\frac{\partial^2 u}{\partial t^2} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right)$$

on the half-plane $y > 0$ with the boundary condition

$$u|_{y=0} = 0 \tag{150}$$

and with any initial conditions:

$$u\,|_{t=0} = \varphi(x, y); \qquad \left.\frac{\partial u}{\partial t}\right|_{t=0} = \varphi_1(x, y)$$

$$(-\infty < x < +\infty;\ y \geqslant 0). \tag{151}$$

We can easily verify that the solution of the problem is given by formula (80) on condition that $\varphi(x, y)$ and $\varphi_1(x, y)$ are continued oddly with respect to the argument y into the interval $(-\infty, 0)$. In fact, with $y = 0$, the first term of (80) can be written as

$$\frac{1}{2\pi a}\int\limits_{x-at}^{x+at}\left[\int\limits_{-\sqrt{a^2t^2-(\alpha-x)^2}}^{+\sqrt{a^2t^2-(\alpha-x)^2}} \frac{\varphi_1(\alpha, \beta)}{\sqrt{a^2t^2-(\alpha-x)^2-\beta^2}}\,d\beta\right]d\alpha$$

and the inner integral is zero for any x and t, since the integrand is an odd function of β. Similarly, the second term of (80) also vanishes, so that condition (150) is in fact satisfied. We might equally have used Fourier's method for the problem, the functions of two variables being represented by Fourier integrals. The proof of the identity of the solution thus obtained with the solution given by (80) presents greater difficulty than in the linear case. The wave equation can be considered in a similar way in the semi-continuum $z \geqslant 0$ with the boundary condition $u = 0$ for $z = 0$. We may also use Fourier's method for solving the wave equation in the unbounded case, when there are only initial conditions. This leads to more complicated working, however, than that carried out above.

§ 18. The equation of telegraphy

181. Fundamental equations. Either of the methods explained above — that of characteristics (d'Alembert's) or that of standing waves (Fourier's) — may also be successfully applied in the case of the *equation of telegraphy*, which has a basic significance in the theory of the propagation of quasi-stationary electric vibrations along cables.

Let us have a circuit made up of out-going and return conductors of length l. We shall assume that the ohmic resistance R, self-inductance L, capacity C and insulation leakage A, all per unit length, are uniformly distributed throughout the circuit, so that the present case differs from that discussed in [I, 181], where we had resistance, self-inductance and capacity lumped at different points of the circuit. Let v and i denote the voltage and current at the section of circuit distant x from the end $x = 0$. These functions of x and t are connected by two differential equations which we shall now derive.

The law of induction says that the voltage drop over an element $\mathrm{d}x$ of the circuit:

$$v-(v+\mathrm{d}v)=-\mathrm{d}v=-\frac{\partial v}{\partial x}\mathrm{d}x$$

is found by adding the ohmic drop $R\,\mathrm{d}x\cdot i$ and the inductive drop $L\,\mathrm{d}x\ \partial i/\partial t$, or in other words, if we divide by $\mathrm{d}x$:

$$\frac{\partial v}{\partial x}+L\frac{\partial i}{\partial t}+Ri=0. \tag{1}$$

Further, the difference between the currents entering and leaving the element $\mathrm{d}x$, i.e.

$$i-(i+\mathrm{d}i)=-\mathrm{d}i=-\frac{\partial i}{\partial x}\mathrm{d}x,$$

is obtained by adding the charging current $C\,\mathrm{d}x\ \partial v/\partial t$ and the leakage current $A\,\mathrm{d}x\cdot v$, so that

$$\frac{\partial i}{\partial x}+C\frac{\partial v}{\partial t}+Av=0. \tag{2}$$

The *boundary conditions* that have to be satisfied at the ends of the circuit are extremely important. If one end of the circuit is open, we must have at this end

$$i=0\quad(\text{for}\ \ x=0\quad\text{or}\quad x=l). \tag{3}$$

In general, if an external electromotive force E, a resistance r and a self-inductance λ are connected to one end of the circuit, we must have at this end:

$$v=E+ri+\lambda\frac{\mathrm{d}i}{\mathrm{d}t}\quad(\text{for}\ \ x=0\quad\text{or}\quad x=l). \tag{4}$$

In particular, if a voltage E only is applied to the end $x=0$, whilst the other end $x=l$ is short-circuited, we have

$$v\,|_{x=0}=E,\qquad v\,|_{x=l}=0. \tag{5}$$

182. Steady state processes. We shall first say a few words about the steady state processes, when the external factors operating on the circuit are either (1) constant or (2) sinusoidal, v and i being taken as independent of t in the first case.

Equations (1) and (2) give us in the first case:

$$\frac{\mathrm{d}v}{\mathrm{d}x}+Ri=0;\quad\frac{\mathrm{d}i}{\mathrm{d}x}+Av=0. \tag{6}$$

On differentiating the first of these equations and making use of the second, we get

$$\frac{d^2v}{dx^2} - RAv = 0. \tag{7}$$

The function v is given at once by the method indicated in [27], and we find

$$v(x) = C_1 e^{bx} + C_2 e^{-bx}, \tag{8}$$

where

$$b = \sqrt{RA}.$$

Having found v, we get i from the first of equations (6):

$$i(x) = -\frac{1}{R}\frac{dv}{dx} = -\frac{b}{R}(C_1 e^{bx} - C_2 e^{-bx}). \tag{9}$$

Examples. **1.** We have conditions (5) in the case of a circuit with a constant voltage E at one end and shorted at the other; hence we can determine the arbitrary constants appearing in equation (8):

$$C_1 + C_2 = E; \quad C_1 e^{bl} + C_2 e^{-bl} = 0,$$

whence

$$C_1 = -\frac{E}{e^{2bl} - 1} = -\frac{Ee^{-bl}}{e^{bl} - e^{-bl}}, \quad C_2 = \frac{Ee^{bl}}{e^{bl} - e^{-bl}},$$

so that we obtain on substituting in (8):

$$v(x) = E\frac{e^{b(l-x)} - e^{-b(l-x)}}{e^{bl} - e^{-bl}} = E\frac{\sinh b(l-x)}{\sinh bl}, \tag{10$_1$}$$

and equation (9) gives

$$i(x) = E\sqrt{\left(\frac{A}{R}\right)}\frac{\cosh b(l-x)}{\sinh lb}. \tag{10$_2$}$$

2. Now let an external sinusoidal electromotive force of given frequency ω act on our circuit; we can convert the physical quantities acting here into vectors, as was done in [I, 180], whilst we shall understand by forced vibrations sinusoidal oscillations of voltage and current in the circuit of the same frequency ω. On recalling the rules of [I, 180] and introducing the current vector $\mathbf{I}$ and voltage vector $\mathbf{V}$, which depend on x in the present case, we can transform the system of differential equations (1) and (2) into:

$$\frac{d\mathbf{V}}{dx} - (R + i\omega L)\mathbf{I} = 0; \quad \frac{d\mathbf{I}}{dx} + (A + i\omega C)\mathbf{V} = 0. \tag{11}$$

On differentiating the first of these equations with respect to x and using the second to eliminate $\mathbf{I}$, we get:

$$\frac{d^2\mathbf{V}}{dx^2} - (R\ i\omega L)(A + i\omega C)\mathbf{V} = 0,$$

whilst precisely the same form of equation is obtained for **I**, as may easily be shown.

Thus, **I** and **V** are solutions of the same second order differential equation. On using the method of [27] and setting

$$(R + i\omega L)(A + i\omega C) = \varkappa^2, \tag{12}$$

we have

$$\mathbf{V} = \mathbf{A}_1 e^{\varkappa x} + \mathbf{A}_2 e^{-\varkappa x}, \tag{13}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are arbitrary constant vectors. On substituting back into the first of equations (11), we find for the vector **I**:

$$\mathbf{I} = -\frac{1}{R + i\omega L}\frac{d\mathbf{V}}{dx} = \sqrt{\frac{A + i\omega C}{R + i\omega L}}\,(\mathbf{A}_2 e^{-\varkappa x} - \mathbf{A}_1 e^{\varkappa x}). \tag{14}$$

The final solution of the problem requires the determination of the constant vectors $\mathbf{A}_1$ and $\mathbf{A}_2$, by making use of the two boundary conditions (of course there is no question here of initial conditions); here, instead of assigning a condition for each end separately, two conditions may be stipulated for the same end, saying by giving the voltage and current vectors there.

However this may be, equations (13) and (14) define the vectors of the forced oscillations which depend on x, i.e. they vary both in amplitude and phase along the circuit. On representing each vector $(m + ni)$ by a point on the complex plane and letting x vary from 0 to l, we get two curves for **V** and **I**, the *vector diagrams of the voltage and current.* It must be recalled when finding the shape of these curves that $\varkappa$ is in general a complex number; on writing

$$\varkappa = a + ib,$$

we have:

$$\mathbf{V} = \mathbf{A}_1 e^{ax}(\cos bx + i \sin bx) + \mathbf{A}_2 e^{-ax}(\cos bx - i \sin bx).$$

Each term on the right-hand side gives a spiral [I, 183], and **V** is obtained by "geometric addition" of these spirals; the radius vector of a point of the curve for **V**, corresponding to a given value of x, is equal to the geometric sum of the radius vectors of the points of the two spirals with the same value of x. We can say the same as regards the vector **I**. On introducing the factor

$$\nu = \sqrt{\frac{R + i\omega L}{A + i\omega C}}. \tag{15}$$

which is known as the *wave impedance*, we can write the expressions for **V** and **I** as

$$\mathbf{V} = \mathbf{A}_1 e^{\varkappa x} + \mathbf{A}_2 e^{-\varkappa x}; \quad \mathbf{I} = \frac{1}{\nu}(\mathbf{A}_2 e^{-\varkappa x} - \mathbf{A}_1 e^{\varkappa x}). \tag{16}$$

If we pass from the vector to the ordinary form, we can write expressions for the required functions v and i of the type:

$$v = V(x) \sin[\omega t + \psi(x)]; \quad i = I(x) \sin[\omega t + \chi(x)], \tag{17}$$

which give the harmonic oscillations of the same frequency ω as the external force, the amplitudes $V(x)$ and $I(x)$, and phases $\psi(x)$ and $\chi(x)$ of these being dependent on the position of the section of the circuit in question.

3. *The circuit with sinusoidal voltage at one end and open at the other end.* Let $\mathbf{V}_0$ denote the voltage vector given at the end $x = 0$. In addition to equations (11), we have the boundary conditions:

$$\mathbf{V}|_{x=0} = \mathbf{V}_0; \mathbf{I}|_{x=l} = 0,$$

which give us, by (16):

$$\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{V}_0; \quad \mathbf{A}_2 e^{-\varkappa l} - \mathbf{A}_1 e^{\varkappa l} = 0.$$

On solving these equations and substituting in (16), we find

$$\mathbf{V} = \mathbf{V}_0 \frac{\cosh \varkappa (l - x)}{\cosh \varkappa l}, \quad \mathbf{I} = \frac{\mathbf{V}_0}{\nu} \frac{\sinh \varkappa (l - x)}{\cosh \varkappa l}.$$

With $x = 0$, we obtain the complex impedance at the point $x = 0$ as

$$\varphi_0 = \nu \frac{\cosh \varkappa l}{\sinh \varkappa l}.$$

183. Transient processes. We propose to compare the two types of forced oscillation, (I) and (II), in the same circuit with the operation of different external factors. The voltage and current of the type (I) oscillation are written v_1, i_1, and those of type (II) by v_2, i_2.

If we suddenly replace the external conditions leading to oscillations (I) by those with which type (II) must be obtained, instead of the transition taking place immediately from (I) to (II), a greater or lesser period of time must elapse, during which *free oscillations* (or *transients*) are excited in the circuit. The transitional period can be infinite in theory but is finite in practice. The transients are characterized by their voltage v and current i, and we shall assume that the transitional state of the circuit is obtained by adding free damped oscillations to state (II), i.e. the voltage and current of the transitional process are defined by the sums

$$v_2 + v; \quad i_2 + i. \tag{18}$$

At the start of the process, $t = 0$, these sums must reduce to v_1 and i_1. The functions v and i must satisfy differential equations (1) and (2) [181] and boundary conditions (3) or (4), depending on the conditions imposed at the ends. Above all, they have to satisfy initial conditions as well, of the form:

$$\left.\begin{aligned} v|_{t=0} &= (v_1 - v_2)|_{t=0} = g(x); \\ i|_{t=0} &= (i_1 - i_2)_{t=0} = \sqrt{\frac{C}{L}}\, h(x).\dagger \end{aligned}\right\} \tag{19}$$

† The factor $\sqrt{C/L}$ is introduced in order to simplify later working.

Instead of seeking v and i directly, we shall express them in terms of a new unknown function w, where we take

$$v = \frac{\partial w}{\partial x}.$$

Equation (2) now gives:

$$\frac{\partial i}{\partial x} + C\frac{\partial^2 w}{\partial x\,\partial t} + A\frac{\partial w}{\partial x} = \frac{\partial}{\partial x}\left(i + C\frac{\partial w}{\partial i} + Aw\right) = 0,$$

whence

$$i + C\frac{\partial w}{\partial t} + Aw = c,$$

where c is independent of x. We can in fact take $c = 0$ without loss of generality, since we can add an arbitrary term not depending on x to w without changing $v = \partial w/\partial x$.

We thus have:

$$v = \frac{\partial w}{\partial x}, \quad i = -C\frac{\partial w}{\partial t} - Aw, \tag{20}$$

and equation (2) is satisfied. On substituting (20) in equation (1), we get an equation which must be satisfied by the function $w(x, t)$, i.e.

$$\frac{\partial}{\partial x}\left(\frac{\partial w}{\partial x}\right) - L\frac{\partial}{\partial t}\left(C\frac{\partial w}{\partial t} + Aw\right) - R\left(C\frac{\partial w}{\partial t} + Aw\right) = 0,$$

or

$$\frac{\partial^2 w}{\partial x^2} - LC\frac{\partial^2 w}{\partial t^2} - (LA + RC)\frac{\partial w}{\partial t} - RAw = 0. \tag{21}$$

This is known as the *equation of telegraphy*.

We simplify this by introducing a new unknown $u(x, t)$ in accordance with the formula:

$$w(x, t) = e^{-\mu t}\, u(x, t), \tag{22}$$

whilst agreeing to choose the constant factor μ in such a way that the terms containing $\partial u/\partial t$ fall out in the equation for u. We have after differentiating and cancelling out $e^{-\mu t}$:

$$\frac{\partial^2 u}{\partial x^2} - LC\left(\mu^2 u - 2\mu\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial t^2}\right) - (LA + RC)\left(-\mu u + \frac{\partial u}{\partial t}\right) - RAu = 0,$$

so that our purpose is achieved if we take

$$2\mu LC - (LA + RC) = 0,$$

i.e.

$$\mu = \frac{LA + RC}{2LC}. \tag{23}$$

If we substitute this value for μ and carry out some simple re-arrangements, we get the equation for u:

$$\frac{\partial^2 u}{\partial t^2} = \frac{1}{LC}\frac{\partial^2 u}{\partial x^2} + \delta^2 u \tag{24}$$

where

$$\delta = \frac{LA - RC}{2LC}.$$

We first distinguish the case when δ can be neglected or is strictly zero, i.e.

$$\frac{R}{L} = \frac{A}{C}. \tag{25}$$

In this case

$$\mu = \frac{R}{L}, \tag{26}$$

and on setting

$$\frac{1}{LC} = a^2, \tag{27}$$

we obtain the equation for u:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, \tag{28}$$

which we investigated above.

The general solution is [164]:

$$u(x, t) = \theta_1(x - at) + \theta_2(x + at), \tag{29}$$

and *the constant* $a = \sqrt{1/LC}$ *gives the velocity of propagation of the excitation along the cable*. Equation (22) gives:

$$w(x, t) = \mathrm{e}^{-\mu t}\left[\theta_1(x - at) + \theta_2(x + at)\right],$$

and finally, we obtain from (20):

$$v(x, t) = \frac{\partial w}{\partial x} = \mathrm{e}^{-\mu t}\left[\theta_1'(x - at) + \theta_2'(x + at)\right],$$

$$\begin{aligned} i(x, t) = -C\frac{\partial w}{\partial t} - Aw = -\mathrm{e}^{-\mu t}\big[-aC\theta_1'(x - at) + aC\theta_2'(x + at) - \\ - \mu C\theta_1(x - at) - \mu C\theta_2(x + at) + A\theta_1(x - at) + A\theta_2(x + at)\big] = \\ = aC\mathrm{e}^{-\mu t}\left[\theta_1'(x - at) - \theta_2'(x + at)\right], \end{aligned}$$

since it is obvious from (25) and (26) that $\mu C = A$, and the remaining terms go out. Instead of the arbitrary functions θ_1 and θ_2, it is more convenient to take directly the functions

$$\varphi_1(x) = \theta_1'(x) \quad \text{and} \quad \varphi_2(x) = \theta_2'(x),$$

after which we obtain the final expressions for v and i:

$$\left.\begin{aligned} v(x,t) &= \mathrm{e}^{-\mu t}\left[\varphi_1(x-at)+\varphi_2(x+at)\right],\\ i(x,t) &= \frac{\mathrm{e}^{-\mu t}}{a}\left[\varphi_1(x-at)-\varphi_2(x+at)\right], \end{aligned}\right\} \tag{30}$$

where we have put $a = \sqrt{L/C}$ for brevity. These are the expressions that we shall utilize. The functions $\varphi_1(x)$ and $\varphi_2(x)$ are determined from the initial conditions (19), which give us:

$$\varphi_1(x)+\varphi_2(x) = g(x)\,; \quad \varphi_1(x)-\varphi_2(x) = h(x)\,,$$

whence

$$\varphi_1(x) = \frac{g(x)+h(x)}{2}\,; \quad \varphi_2(x) = \frac{g(x)-h(x)}{2}\,. \tag{31}$$

We could reckon the problem solved if the functions $g(x)$ and $h(x)$, or what amounts to the same thing, $\varphi_1(x)$ and $\varphi_2(x)$, were specified throughout the interval $(-\infty, +\infty)$; but in fact, we only know these in the interval $(0, l)$, and in order to be able to make use of the solution obtained we have to *continue them outside this interval*. We can do this with the aid of the boundary conditions as in the case of a string, and *the physical meaning of the continuation* also *amounts* here *to a wave reflection occurring at the ends of the circuit.*

The phenomena corresponding to solution (30) are analogous to those worked out above for a string. We have two waves here, the direct and the reverse, which are reflected on arrival at the ends. The essential difference between this case and that of a string is the presence of the factor $\mathrm{e}^{-\mu t}$, which diminishes with time and produces the damping of the oscillation; the greater the exponent μ *(the logarithmic decrement of the damping)*, the faster the damping.

184. Examples. *If the end $x = l$ is open*, the condition

$$i\,|_{x=l} = 0$$

gives us by (30):

$$\varphi_2(l+at) = \varphi_1(l-at)$$

or on replacing at by x:

$$\varphi_2(l+x) = \varphi_1(l-x)\,,$$

i.e. *the wave is reflected without change of magnitude or sign* at this end, since $\varphi_2(x)$ is an even continuation of $\varphi_1(x)$. The same is evidently true if the open end is at $x = 0$.

If the end $x = l$ is short-circuited, i.e.

$$v\,|_{x=l} = 0\,,$$

we obtain on taking (30) into account and replacing *at* by x:

$$\varphi_2(l+x) = -\varphi_1(l-x),$$

i.e. *the wave is reflected with a change of sign but with the same absolute value,* since $\varphi_2(x)$ is an odd continuation of $\varphi_1(x)$. The further continuation is obtained as in the case of a string.

1. The circuit open at one end has a harmonic alternating current of frequency ω switched into it. The harmonic oscillations deduced above [182] correspond to the final steady state (II):

$$v_2 = V(x)\sin[\omega t + \psi(x)]; \quad i_2 = I(x)\sin[\omega t + \chi(x)].$$

If the circuit was empty before the inclusion of the current, we have:

$$v_1 = 0, \quad i_1 = 0.$$

Hence the initial conditions become, by (19):

$$v|_{t=0} = -V(x)\sin\psi(x) = g(x),$$
$$i|_{t=0} = -I(x)\sin\chi(x) = \frac{1}{a}h(x).$$

The boundary conditions are as follows. At the end $x = l$ we must have:

$$i|_{x=l} = 0.$$

We can take at the end $x = 0$:

$$v|_{x=0} = 0,$$

since in the present transient process we are only interested in the oscillations that are due to the difference of the initial conditions of the circuit from the

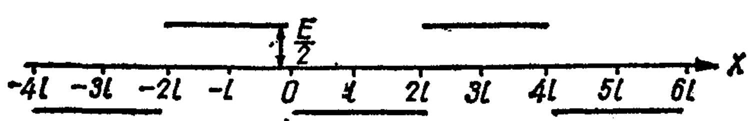

FIG. 138

forced oscillations of frequency ω. We define $\varphi_1(x)$ and $\varphi_2(x)$ in accordance with (31) and then continue these, in an odd manner past the end $x = l$, and evenly past $x = 0$.

2. We consider the damped process due to the initial conditions

$$v|_{t=0} = -E; \quad i|_{t=0} = 0,$$

where E is a constant, and with the boundary conditions

$$v|_{x=0} = 0; \quad i|_{x=l} = 0.$$

Equations (31) give

$$\varphi_1(x) = \varphi_2(x) = -\frac{1}{2}E \quad \text{where} \quad 0 < x < l,$$

and we find from the boundary conditions:

$$\varphi_1(-x) = -\varphi_2(x); \quad \varphi_1(l-x) = \varphi_2(l+x), \tag{32}$$

from which it is clear that $\varphi_2(x)$ is an even continuation of $\varphi_1(x)$ in the interval $(l, 2l)$, whilst $\varphi_1(x)$ is an odd continuation of $\varphi_2(x)$ in the interval $(-l, 0)$, i.e.

$$\varphi_2(x) = -\frac{E}{2} \qquad \text{for} \qquad 0 < x < 2l$$

$$\varphi_1(x) = \begin{cases} \frac{1}{2}E & \text{for} \quad -l < x < 0 \\ -\frac{1}{2}E & \text{for} \quad 0 < x < l. \end{cases}$$

On writing $(l + x)$ for x in the second of equations (32) and comparing the equation obtained with the first of (32), we get:

$$\varphi_2(2l + x) = -\varphi_2(x),$$

and similarly,

$$\varphi_1(2l - x) = -\varphi_1(-x),$$

i.e. $\varphi_1(x)$ and $\varphi_2(x)$ change sign on addition of $2l$ to the argument, so that their period is $4l$.

On putting together everything that has been said, we see that $\varphi_1(x)$ and $\varphi_2(x)$ are in fact the same, their graph being illustrated in Fig. 138.

We obtain v and i by moving this graph to the left and right with velocity a; v is half the sum of the ordinates multiplied by $e^{-\mu t}$, and i is half the difference of the ordinates multiplied by $e^{-\mu t}/a$.

The graph of the voltage at the end $x = l$ is shown in Fig. 139, the steady state $v_2 = E$ being added to the free oscillation v. The $\tau = 4l/a$ denotes the period of the free oscillation.

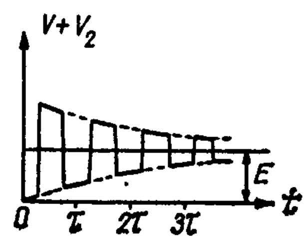

FIG. 139

If we have ohmic resistance r_l, self-inductance λ_l and capacity γ_l at the end $x = l$, conditions (4) give the following relationship for continuation of $\varphi_2(x)$ in $(l, 2l)$:

$$e^{-\mu t}\left[\varphi_1(l - at) + \varphi_2(l + at)\right] = \left[r_l + \lambda_l \frac{d}{dt}\right]\left\{\frac{e^{-\mu t}}{a}\left[\varphi_1(l-at) - \varphi_2(l+at)\right]\right\}. \quad (33)$$

On replacing the argument at by x, this leads to a differential equation for the unknown function:

$$\Phi(x) = \varphi_2(l + x) \qquad 0 < x < l.$$

We get a similar result for the continuation of $\varphi_1(x)$ in $(-l, 0)$ by applying the boundary condition at the end $x = 0$.

3. The circuit is terminated at $x = l$ by the pure ohmic resistance r_l. Equation (33) now becomes:

$$e^{-\mu t}\left[\varphi_1(l - at) + \varphi_2(l + at)\right] = r_l \frac{e^{-\mu t}}{a}\left[\varphi_1(l - at) - \varphi_2(l + at)\right],$$

whence we find $\varphi_2(l+x)$ after writing x for at:

$$\varphi_2(l+x) = q\varphi_1(l-x), \qquad q = \frac{r_l - a}{r_l + a}. \tag{34}$$

The wave is thus multiplied by the factor q on reflection at the end $x = l$ in the present case. Obviously, $|q| \leqslant 1$, i.e. the absolute value of the wave diminishes or remains constant, with *absorption* occurring in the former case. The factor vanishes $r_l = a$, and *total absorption of the wave* takes place. With $r_l = \infty$, $q = 1$ and we get reflection of the wave without change, which is obvious since this case is equivalent to the open circuit case.

Having thus continued $\varphi_2(x)$ into $(l, 2l)$, and correspondingly $\varphi_1(x)$ into $(-l, 0)$, we can continue $\varphi_2(x)$ into $(2l, 3l)$ in accordance with (34), and so on.

Of course we no longer obtain a periodic function in this case, and assuming $|q| < 1$, subsequent reflections lead to stronger and stronger absorptions. The function $\varphi_2(x)$ is now defined for $x > 0$, and $\varphi_1(x)$ for $x < l$; but this is just what we need, since the arguments $(l - at)$ and $(l + at)$, on which $\varphi_1(x)$ and $\varphi_2(x)$ depend, in fact satisfy these inequalities.

185. Generalized equation of vibration of a string. We have considered the equation in the particular case $\delta = 0$. Before turning to the general case, we investigate theoretically the generalized wave equation in the linear case:

$$\frac{\partial^2 v}{\partial t^2} = a^2 \frac{\partial^2 v}{\partial x^2} + a_1 \frac{\partial v}{\partial x} + a_2 \frac{\partial v}{\partial t} + a_3 v, \tag{35}$$

where we take the first coefficient a^2 as positive and the remainder of arbitrary sign. We replace v by a new required function u in accordance with

$$v = e^{\alpha t + \beta x}\, u \tag{36}$$

and show as above that an α and β can always be chosen so that terms containing first order partial derivatives fall out in the equation for u. We substitute (36) in equation (35), cancel out $e^{\alpha t + \beta x}$ and collect like terms, and arrive at the equation

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} + (a_1 + 2a^2\beta)\frac{\partial u}{\partial x} + (a_2 - 2\alpha)\frac{\partial u}{\partial t} +$$
$$+ (a_3 + a^2\beta^2 + a_1\beta + a_2\alpha - \alpha^2)\, u,$$

or, after setting $\alpha = a_2/2$, $\beta = -a_1/2a^2$:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} + c^2 u, \tag{37}$$

where the coefficient c^2 can be either positive or negative, i.e. c is to be reckoned either positive or pure imaginary.

We shall solve equation (37) for an infinite x axis with the initial conditions:

$$u\Big|_{t=0} = 0;\ \frac{\partial u}{\partial t}\Big|_{t=0} = \omega(x). \tag{38}$$

Instead of solving directly the problem posed by (37) and (38), we attack the problem defined by the following equation and initial conditions:

$$\frac{\partial^2 w}{\partial t^2} = a^2\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right), \tag{39$_1$}$$

$$w\Big|_{t=0} = 0;\quad \frac{\partial w}{\partial t}\Big|_{t=0} = \omega(x)\,\mathrm{e}^{cy/a}. \tag{39$_2$}$$

We can write down the solution of this immediately on using expression (80) of [172]:

$$w(x, y, t) = \frac{1}{2\pi a}\iint\limits_{C_{at}} \frac{\omega(\alpha)\,\mathrm{e}^{\beta c/a}\,\mathrm{d}\alpha\,\mathrm{d}\beta}{\sqrt{a^2t^2 - (\alpha - x)^2 - (\beta - y)^2}},$$

where C_{at} is the circle with radius at and centre (x, y). On introducing new variables $\alpha' = \alpha - x$ and $\beta' = \beta - y$ instead of α and β, we can transform the double integral to that over the circle C'_{at} with radius a and centre at the origin:

$$w(x, y, t) = \frac{1}{2\pi a}\iint\limits_{C'_{at}} \frac{\omega(\alpha' + x)\,\mathrm{e}^{c(\beta'+y)/a}\,\mathrm{d}\alpha'\,\mathrm{d}\beta'}{\sqrt{a^2t^2 - \alpha'^2 - \beta'^2}}$$

or alternatively, after taking $\mathrm{e}^{cy/a}$ outside the integral sign, we can write

$$w(x, y, t) = \mathrm{e}^{cy/a}\,u(x, t), \tag{40}$$

where the factor

$$u(x, t) = \frac{1}{2\pi a}\iint\limits_{C'_{at}} \frac{\omega(\alpha' + x)\,\mathrm{e}^{c\beta'/a}\,\mathrm{d}\alpha'\,\mathrm{d}\beta'}{\sqrt{a^2t^2 - \alpha'^2 - \beta'^2}} \tag{41}$$

is clearly independent of y. We show that (41) in fact solves our original problem, i.e. satisfies equation (37) and initial conditions (38). The function w satisfies equation (39_1), and on substituting expression (40) in (39_1), we obtain equation (37) for u after cancelling out $\mathrm{e}^{cy/a}$. The initial conditions for u are obtained at once from initial conditions (39_2) for w and expression (40). Thus (41) gives the solution of equation (37) with initial conditions (38). We shall write the expression on the right-hand side of (41) in a new form.

We reduce the double integral over the circle C'_{at} to two quadratures:

$$u(x,t) = \\ = \frac{1}{2\pi a} \int_{-at}^{+at} \left[\int_{-\sqrt{a^2t^2-\alpha'^2}}^{+\sqrt{a^2t^2-\alpha'^2}} \frac{e^{c\beta'/a}}{\sqrt{a^2t^2 - \alpha'^2 - \beta'^2}} d\beta' \right] \omega(\alpha' + x)\, d\alpha'. \tag{42}$$

We replace the variable of integration β' of the inner integral by a new variable φ, in accordance with $\beta' = \sqrt{a^2 t^2 - \alpha'^2} \sin \varphi$, which gives us the integral in the form:

$$\int_{-\pi/2}^{+\pi/2} e^{c/a \sqrt{a^2t^2-\alpha'^2} \sin\varphi} d\varphi$$

On introducing a new transcendental function $I(z)$, defined by an integral and depending on the parameter z:

$$I(z) = \frac{1}{\pi} \int_{-\pi/2}^{+\pi/2} e^{z \sin\varphi} d\varphi, \tag{43}$$

we can write (42) in the form:

$$u(x,t) = \frac{1}{2a} \int_{-at}^{+at} I\left(\frac{c}{a} \sqrt{a^2t^2 - \alpha'^2}\right) \omega(\alpha' + x)\, d\alpha',$$

or, on introducing as variable of integration $\alpha = \alpha' + x$:

$$u(x,t) = \frac{1}{2a} \int_{x-at}^{x+at} I\left(\frac{c}{a} \sqrt{a^2t^2 - (\alpha - x)^2}\right) \omega(\alpha)\, d\alpha.$$

We differentiate the solution obtained with respect to t and obtain as in [171] a new solution $u_1 = \partial u/\partial t$ of equation (37), no longer satisfying initial conditions (38) but instead the conditions

$$u\Big|_{t=0} = \omega(x); \quad \frac{\partial u}{\partial t}\Big|_{t=0} = 0. \tag{44}$$

The solution of (37) satisfying the general initial conditions

$$u\Big|_{t=0} = \varphi(x); \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \varphi_1(x), \tag{45}$$

is simply obtained by taking $\omega(x) = \varphi_1(x)$ in initial conditions (38) and $\omega(x) = \varphi(x)$ in conditions (44) then adding the respective expressions

for u; this gives us

$$u(x,t) = \frac{1}{2a}\int\limits_{x-at}^{x+at} I\left(\frac{c}{a}\sqrt{a^2t^2-(\alpha-x)^2}\right)\varphi_1(\alpha)\,d\alpha +$$

$$+\frac{\partial}{\partial t}\left[\frac{1}{2a}\int\limits_{x-at}^{x+at} I\left(\frac{c}{a}\sqrt{a^2t^2-(\alpha-x)^2}\right)\varphi(\alpha)\,d\alpha\right]. \qquad (46)$$

On carrying out the differentiation with respect to t both for the upper and lower limits and under the integral sign, and noting that $I(0) = 1$ by (43), we can rewrite (46) in the form:

$$u(x,t) = \frac{\varphi(x-at)+\varphi(x+at)}{2} +$$

$$+\frac{1}{2a}\int\limits_{x-at}^{x+at} I\left(\frac{c}{a}\sqrt{a^2t^2-(\alpha-x)^2}\right)\varphi_1(\alpha)\,d\alpha +$$

$$+\frac{ct}{2}\int\limits_{x-at}^{x+at}\frac{1}{\sqrt{a^2t^2-(\alpha-x)^2}}\,I'\left(\frac{c}{a}\sqrt{a^2t^2-(\alpha-x)^2}\right)\varphi(\alpha)\,d\alpha, \qquad (47)$$

where $I'(z)$ denotes the derivative of $I(z)$ with respect to the argument z.

We now establish a relationship between the function $I(z)$ and the zero order Bessel function [48]:

$$J_0(x) = \sum_{s=0}^{\infty}\frac{(-1)^s}{(s!)^2}\left(\frac{x}{2}\right)^{2s}. \qquad (48)$$

On expanding $e^{z\sin\varphi}$ in a power series:

$$e^{z\sin\varphi} = \sum_{n=0}^{\infty}\frac{z^n\sin^n\varphi}{n!},$$

then integrating term by term over the interval $(-\pi/2, +\pi/2)$, as is possible in view of the uniform convergence of the series, we get

$$I(z) = \sum_{n=0}^{\infty}\frac{z^n}{n!}\cdot\frac{1}{\pi}\int\limits_{-\pi/2}^{+\pi/2}\sin^n\varphi\,d\varphi.$$

The integrals on the right evidently vanish with n odd, whilst we have for even $n = 2s$ [I, 100]:

$$\int\limits_{-\pi/2}^{+\pi/2}\sin^{2s}\varphi\,d\varphi = 2\int\limits_{0}^{\pi/2}\sin^{2s}\varphi\,d\varphi = \frac{(2s-1)(2s-3)\ldots 1}{2s\cdot(2s-2)\ldots 2}\,\pi,$$

whence it follows that:

$$I(z) = \sum_{s=0}^{\infty} \frac{z^{2s}}{(2s)!} \cdot \frac{(2s-1)(2s-3)\dots 1}{2s \cdot (2s-2)\dots 2},$$

or

$$I(z) = \sum_{s=0}^{\infty} \frac{1}{(s!)^2} \left(\frac{z}{2}\right)^{2s}. \tag{49}$$

On comparing this expansion with (48), we see that

$$I(z) = J_0(iz). \tag{50}$$

186. The general case of an infinite circuit. We shall now discuss the equation of telegraphy for an infinite circuit. A preliminary observation concerns equation (21), obtained for the auxiliary function w in [183]: this is also the equation which must be satisfied individually by the voltage v and current i.

We return to the fundamental equations (1) and (2) and eliminate i. We do this by differentiating equation (1) with respect to x then replacing $\partial i/\partial x$ by its expression from equation (2):

$$\frac{\partial^2 v}{\partial x^2} + L \frac{\partial^2 i}{\partial t\, \partial x} + R \frac{\partial i}{\partial x} = 0,$$

i.e.

$$\frac{\partial^2 v}{\partial x^2} - L \frac{\partial}{\partial t}\left(C \frac{\partial v}{\partial t} + Av\right) - R\left(C \frac{\partial v}{\partial t} + Av\right) = 0$$

whence we have equation (21) for v:

$$\frac{\partial^2 v}{\partial x^2} - LC \frac{\partial^2 v}{\partial t^2} - (LA + RC)\frac{\partial v}{\partial t} - RAv = 0. \tag{51}$$

If we had eliminated the voltage v from equations (1) and (2), we should have again arrived at this equation as the equation for i.

Having determined v, we can find i so that it satisfies equations (1) and (2). For instance, use of equation (2) gives us:

$$i = -\int \left(C \frac{\partial v}{\partial t} + Av\right) \mathrm{d}x + B(t), \tag{52}$$

where the integration is with respect to x with constant t and $B(t)$ is at present an arbitrary function of t. We set this expression for i in equation (1) and differentiate with respect to the parameter t under the integral sign:

$$\frac{\partial v}{\partial x} - \int \left(LC \frac{\partial^2 v}{\partial t^2} + LA \frac{\partial v}{\partial t}\right) \mathrm{d}x - \int \left(RC \frac{\partial v}{\partial t} + RAv\right) \mathrm{d}x + $$
$$+ LB'(t) + RB(t) = 0. \tag{53}$$

Differentiation with respect to x of the sum of the first three terms gives us zero by (51), i.e. the sum is a known function of t only, and we obtain a linear first order equation for $B(t)$. The arbitrary constant obtained by integrating this latter equation is in general defined by the initial conditions.

As above in [183], equation (51) reduces to the form

$$\frac{\partial^2 u}{\partial t^2} = \frac{1}{LC}\frac{\partial^2 u}{\partial x^2} + c^2 u \tag{54}$$

with the aid of the substitution

$$v(x, t) = e^{-\mu t} u(x, t), \tag{55}$$

where

$$\mu = \frac{LA + RC}{2LC}, \qquad c = \frac{|LA - RC|}{2LC}. \tag{56}$$

If we are given v and i along the circuit at $t = 0$, we must also know $\partial v/\partial x$ and $\partial i/\partial x$ at $t = 0$, whilst equations (1) and (2) give $\partial v/\partial t$ and $\partial i/\partial t$ at $t = 0$. Hence we can suppose that, along with equation (51), we have the general initial conditions:

$$v\Big|_{t=0} = \Phi(x); \quad \frac{\partial v}{\partial t}\Big|_{t=0} = \Psi(x). \tag{57}$$

We make use of (55) to obtain the following initial conditions for u:

$$u\Big|_{t=0} = \Phi(x); \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \mu\Phi(x) + \Psi(x). \tag{58}$$

We finally get, on applying equation (47) for u and taking (55) into account:

$$\begin{aligned} v(x, t) = \frac{1}{2} e^{-\mu t} \Big\{ &\Phi(x - at) + \Phi(x + at) + \\ &+ \frac{1}{a}\int_{x-at}^{x+at} [\mu\Phi(\alpha) + \Psi(\alpha)]\, I\left(\frac{c}{a}\sqrt{a^2t^2 - (\alpha - x)^2}\right) d\alpha + \\ &+ \frac{ct}{2}\int_{x-at}^{x+at} \frac{1}{\sqrt{a^2t^2 - (\alpha - x)^2}}\, I'\left(\frac{c}{a}\sqrt{a^2t^2 - (\alpha - x)^2}\right)\Phi(\alpha)\, d\alpha \Big\}, \end{aligned} \tag{59}$$

where μ and c are as given above, and $a = 1/\sqrt{LC}$.

We have here a definite velocity a of propagation of the excitation, as in the case of a vibrating string; thus, if the functions $\Phi(x)$ and $\Psi(x)$ giving the initial excitation differ from zero only in a finite interval $p \leqslant x \leqslant q$ and we apply (57) at the point x vhere $x > q$, we must have $v(x, t)$ zero up to the instant $t = (x - q)/a$. The essential

difference from the case of a string lies in the fact that $v(x, t)$ does not vanish or become constant after passage of the rear front of the initial excitation, but is still a function of x and t. In fact, if $t > (x - p)/a$, the terms outside the integral signs in (59) vanish, whereas the integrals remain, the interval of integration being the constant (p, q). Nevertheless, the variables x and t appear as parameters under the integral signs.

If, for instance, current is absent in the circuit at $t = 0$, whilst the potential v is given by $\Phi(x)$, we have by equation (2):

$$\frac{\partial v}{\partial t}\bigg|_{t=0} = -\frac{A}{C}\Phi(x). \tag{60}$$

If we take $A = 0$, i.e. neglect the leakage, the right-hand side is zero.

187. Fourier's method for a finite circuit. Fourier's method may readily be used for integrating equation (51) with given initial and boundary conditions in the case of a finite circuit. Let the end $x = 0$ be subjected to a given constant voltage E, whilst $v = 0$ at the other end, so that the boundary conditions are

$$u\,|_{x=0} = E, \qquad v\,|_{x=l} = 0. \tag{61}$$

Furthermore, let there be neither voltage nor current in the circuit at the initial instant $t = 0$, i.e.

$$v\,|_{t=0} = 0, \qquad i\,|_{t=0} = 0 \tag{62}$$

for $0 < x < l$.

Equations (1) and (2) show us that here:

$$\frac{\partial v}{\partial t}\bigg|_{t=0} = 0, \qquad \frac{\partial i}{\partial t}\bigg|_{t=0} = 0. \tag{63}$$

We thus have to integrate equation (51) with boundary conditions (61) and with initial conditions

$$v\,|_{t=0} = 0;\ \frac{\partial v}{\partial t}\bigg|_{t=0} = 0 \qquad (0 < x < l). \tag{64}$$

We start by constructing a solution of (51), $v = F(x)$, which depends only on x and which would satisfy boundary conditions (61). We obtain as an equation for $F(x)$:

$$F''(x) - b^2F(x) = 0 \qquad (b^2 = RA).$$

We actually found the solution of this equation in the example of [182], which also satisfied conditions (61):

$$F(x) = E\frac{\sinh b(l-x)}{\sinh bl}. \tag{65}$$

We now introduce a new required function $w(x, t)$ in place of $v(x, t)$, in accordance with

$$w(x,t) = v(x,t) - F(x). \tag{66}$$

We have the same equation (51) for $w(x, t)$, together with the homogeneous boundary conditions

$$w\big|_{x=0} = 0; \quad w\big|_{x=l} = 0 \tag{67}$$

and the initial conditions

$$w\big|_{t=0} = -F(x); \quad \frac{\partial w}{\partial t}\bigg|_{t=0} = 0. \tag{68}$$

We shall simplify the writing by putting equation (51) for w in the form:

$$\frac{\partial^2 w}{\partial x^2} - a^2\frac{\partial^2 w}{\partial t^2} - 2h\frac{\partial w}{\partial t} - b^2 w = 0, \tag{69}$$

where

$$a^2 = LC; \quad 2h = LA + RC; \quad b^2 = RA. \tag{70}$$

We now proceed as usual with Fourier's method. We seek a solution of (69) as the product of a function of x only and a function of t only:

$$w = XT.$$

We obtain on substituting in (69) and separating the variables:

$$\frac{X''}{X} = \frac{a^2T'' + 2hT' + b^2T}{T} = -\frac{m^2\pi^2}{l^2},$$

where m^2 is at present an arbitrary constant. We have two linear equations with constant coefficients:

$$X'' + \frac{m^2\pi^2}{l^2}X = 0$$

$$a^2T'' + 2hT' + \left(b^2 + \frac{m^2\pi^2}{l^2}\right)T = 0.$$

We now take the boundary conditions (67) into account and select the solutions

$$X_m = \sin\frac{m\pi x}{l} \qquad (m = 1, 2, \ldots)$$

of the first equation, where m is a positive integer. The equation for T has the general solution

$$T_m = A_m e^{\alpha_m t} + A'_m e^{\alpha'_m t},$$

where A_m and A'_m are arbitrary constants, whilst α_m and α'_m are the roots of the equation

$$a^2 l^2 \alpha^2 + 2hl^2 \alpha + (b^2 l^2 + m^2 \pi^2) = 0, \tag{71}$$

the circuit constants R, L, C and A being assumed to be such that this equation has different roots for any integral m. We thus obtain an infinite set of solutions satisfying the boundary conditions:

$$w_m = (A_m e^{\alpha_m t} + A'_m e^{\alpha'_m t}) \sin \frac{m\pi x}{l} \tag{72}$$

We take the sum of these solutions:

$$w = \sum_{m=1}^{\infty} (A_m e^{\alpha_m t} + A'_m e^{\alpha'_m t}) \sin \frac{m\pi x}{l} \tag{73}$$

and select the constants A_m and A'_m so as to satisfy the initial conditions (68). This gives us:

$$\left.\begin{aligned} &\sum_{m=1}^{\infty} (A_m + A'_m) \sin \frac{m\pi x}{l} = -F(x) \\ &\sum_{m=1}^{\infty} (\alpha_m A_m + \alpha'_m A'_m) \sin \frac{m\pi x}{l} = 0. \end{aligned}\right\} \quad (0 < x < l)$$

Having found the Fourier coefficients in the ordinary way, we get two equations for A_m and A'_m:

$$\left.\begin{aligned} &A_m + A'_m = -\frac{2}{l}\int_0^l F(x) \sin \frac{m\pi x}{l}\, dx \\ &\alpha_m A_m + \alpha'_m A'_m = 0. \end{aligned}\right\} \tag{74}$$

Having substituted for $F(x)$ from (65), we can carry out the integration and obtain:

$$\frac{2}{l}\int_0^l F(x) \sin \frac{m\pi x}{l}\, dx = \frac{2m\pi}{b^2 l^2 + m^2 \pi^2} E.$$

We have after solving simultaneous equations (74):

$$A_m = \frac{2m\pi}{b^2l^2 + m^2\pi^2} \cdot E \frac{\alpha'_m}{\alpha_m - \alpha'_m}; \quad A'_m = - \frac{2m\pi}{b^2l^2 + m^2\pi^2} \cdot E \frac{\alpha_m}{\alpha_m - \alpha'_m}.$$

On substituting this in expression (73), we get:

$$w = E \sum_{m=1}^{\infty} \frac{2m\pi}{b^2l^2 + m^2\pi^2} \cdot \frac{\alpha'_m e^{\alpha_m t} - \alpha_m e^{\alpha'_m t}}{\alpha_m - \alpha'_m} \sin \frac{m\pi x}{l}. \tag{75}$$

The roots of equation (71) are either real and negative, or complex conjugates with negative real parts. In either case, solution (75) is damped with increasing t. It defines the transient process from the empty circuit to the steady state defined by function (65). Equation (66) gives us the final expression for the voltage:

$$v = E \frac{\sinh b(l-x)}{\sinh bl} + E \sum_{m=0}^{\infty} \frac{2m\pi}{b^2l^2 + m^2\pi^2} \cdot \frac{\alpha'_m e^{\alpha_m t} - \alpha_m e^{\alpha'_m t}}{\alpha_m - \alpha'_m} \sin \frac{m\pi x}{l}. \tag{76}$$

Solution of quadratic equation (71) gives us roots of the form

$$\alpha_m = -\nu + k_m, \quad \alpha'_m = -\nu - k_m, \tag{77}$$

where

$$\nu = \frac{h}{a^2}; \quad k_m = \frac{1}{a^2 l} \sqrt{h^2l^2 - a^2(b^2l^2 + m^2\pi^2)}. \tag{78}$$

On substituting in (76), we can rewrite this as

$$v = E \frac{\sinh b(l-x)}{\sinh bl} -$$

$$- E e^{-\nu t} \sum_{m=1}^{\infty} \frac{2m\pi}{b^2l^2 + m^2\pi^2} \left(\cosh k_m t + \frac{\nu}{k_m} \sinh k_m t\right) \sin \frac{m\pi x}{l}. \tag{79}$$

We now find i by the method of the previous section. Equation (2) gives us:

$$\frac{\partial i}{\partial x} = - AE \frac{\sinh b(l-x)}{\sinh bl} +$$

$$+ AE e^{-\nu t} \sum_{m=1}^{\infty} \frac{2m\pi}{b^2l^2 + m^2\pi^2} \left(\cosh k_m t + \frac{\nu}{k_m} \sinh k_m t\right) \sin \frac{m\pi x}{l} +$$

$$+ CE e^{-\nu t} \sum_{m=1}^{\infty} \frac{2m\pi}{b^2l^2 + m^2\pi^2} \left(k_m - \frac{\nu^2}{k_m}\right) \sinh k_m t \sin \frac{m\pi x}{l},$$

or, on observing that, by (78):

$$\nu^2 - k_m^2 = \frac{b^2l^2 + m^2\pi^2}{a^2l^2},$$

we can integrate with respect to x and use the fact that $a^2 = LC$:

$$i = \frac{AE}{b} \cdot \frac{\cosh b(l-x)}{\sinh bl} -$$

$$- 2AEle^{-\nu t} \sum_{m=1}^{\infty} \frac{1}{b^2l^2 + m^2\pi^2} \left(\cosh k_m t + \frac{\nu}{k_m} \sinh k_m t\right) \cos \frac{m\pi x}{l} +$$

$$+ \frac{2E}{Ll} e^{-\nu t} \sum_{m=1}^{\infty} \frac{1}{k_m} \sinh k_m t \cos \frac{m\pi x}{l} + B(t). \tag{80}$$

We find an equation for $B(t)$ on substituting in equation (1):

$$LB'(t) + RB(t) = 0,$$

whence

$$B(t) = B_0 e^{-Rt/L}, \tag{81}$$

where B_0 is an arbitrary constant which has to be determined from the condition that i is zero throughout the length of the circuit at $t = 0$. We substitute expression (81) in (80) and then set $t = 0$ and $i = 0$, which gives us

$$0 = \frac{AE}{b} \frac{\cosh b(l-x)}{\sinh bl} - 2AEl \sum_{m=1}^{\infty} \frac{1}{b^2l^2 + m^2\pi^2} \cos \frac{m\pi x}{l} + B_0. \tag{82}$$

But the cosine Fourier expansion of the first term on the right in the interval $0 < x < l$ is

$$\frac{AE}{b} \frac{\cosh b(l-x)}{\sinh bl} = \frac{AE}{lb^2} + 2AEl \sum_{m=1}^{\infty} \frac{1}{b^2l^2 + m^2\pi^2} \cos \frac{m\pi x}{l} \quad (0 < x < l),$$

and condition (82) gives:

$$B_0 - \frac{AE}{lb^2} = - \frac{E}{Rl},$$

so that

$$B(t) = - \frac{E}{Rl} e^{-Rt/L}.$$

Substitution of this expression for $B(t)$ in (80) gives us the final expression for the current.

A detailed discussion of the above method of solution can be found in A. V. Krylov's article "The propagation of currents along cables" (*Zhurnal prikladnoi fiziki*, Vol. VI, sec. 2, p. 66, 1929).

188. The generalized wave equation. We considered the generalized wave equation in the linear case in [185], i.e. with two independent variables. We can use the same method for the generalized wave

equation with three or four independent variables. We shall take the velocity $a = 1$ in the equation, so as to simplify later expressions; these can be turned into formulae for any a simply by replacing t in them by at.

We take the equation for the infinite plane:

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + c^2 u \tag{83}$$

with the initial conditions

$$u\Big|_{t=0} = 0\,; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \omega(x, y). \tag{84}$$

We firstly consider a new problem, that of integrating the wave equation

$$\frac{\partial^2 w}{\partial t^2} = \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}$$

with initial conditions

$$w\Big|_{t=0} = 0\,; \quad \frac{\partial w}{\partial t}\Big|_{t=0} = \omega(x, y)\,e^{cz}.$$

The new problem is solved at once by Poisson's formula:

$$w = \frac{t}{4\pi}\int_0^{2\pi}\int_0^{\pi} \omega(x + t\sin\theta\cos\varphi,\ y + t\sin\theta\sin\varphi)\,e^{c(z + t\cos\theta)}\sin\theta\,d\theta\,d\varphi.$$

We can rewrite this in the form:

$$w(x, y, z, t) = e^{cz} u(x, y, t),$$

where

$$u(x, y, t) =$$
$$= \frac{t}{4\pi}\int_0^{2\pi}\int_0^{\pi} \omega(x + t\sin\theta\cos\varphi,\ y + t\sin\theta\sin\varphi)\,e^{ct\cos\theta}\sin\theta\,d\theta\,d\varphi, \tag{85}$$

It can be shown, precisely as in [185], that this latter function satisfies equation (83) and initial conditions (84). We now simplify (85) by replacing θ with a new variable of integration ϱ in accordance with $t\cos\theta = \varrho$, from which we have:

$$t\sin\theta\,d\theta = -\,d\varrho, \qquad \sin\theta = \sqrt{1 - \frac{\varrho^2}{t^2}}.$$

The integral over θ in (85) becomes with the new variable:

$$\frac{1}{t}\int_{-t}^{+t} \omega\left(x + \sqrt{t^2-\varrho^2}\cos\varphi,\, y + \sqrt{t^2-\varrho^2}\sin\varphi\right) e^{c\varrho}\, d\varrho\,,$$

which may be written, if we split the interval of integration into $(-t, 0)$ and $(0, t)$ and replace ϱ by $(-\varrho)$ in the first sub-interval:

$$\frac{2}{t}\int_{0}^{t} \omega\,(x + \sqrt{t^2-\varrho^2}\cos\varphi, y + \sqrt{t^2-\varrho^2}\sin\varphi)\cosh c\varrho\, d\varrho\,,$$

so that (85) may be written as

$$u\,(x,\, y,\, t) =$$
$$= \int_0^t \left[\frac{1}{2\pi}\int_0^{2\pi} \omega\,(x + \sqrt{t^2-\varrho^2}\cos\varphi,\; y + \sqrt{t^2-\varrho^2}\, sin\,\varphi)\, d\varphi\right]\cosh c\varrho\, d\varrho\,.$$

The integration with respect to φ in this expression gives the arithmetic mean of the function $\omega(x, y)$ over the circle in the xy plane with centre (x, y) and radius $\sqrt{t^2-\varrho^2}$. On denoting this arithmetic mean as $T_{\sqrt{t^2-\varrho^2}}\,\{\omega(x, y)\}$, we can write (85) in the final form

$$u\,(x,\, y,\, t) = \int_0^t T_{\sqrt{t^2-\varrho^2}}\left\{\omega\,(x,\, y)\right\}\cosh c\varrho\, d\varrho\,. \tag{86}$$

We note that if $c = c_1\, i$ is pure imaginary, $\cosh c\varrho = \cos c_1\varrho$. On differentiating our solution with respect to t, we get the solution $u_1 = \partial u/\partial t$ of equation (83) satisfying the initial conditions:

$$u_1\Big|_{t=0} = \omega\,(x,\, y)\,; \quad \frac{\partial u_1}{\partial t}\Big|_{t=0} = 0\,.$$

Similarly, the integration of the equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + c^2 u \tag{87}$$

with the initial conditions

$$u\Big|_{t=0} = 0\,; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \omega\,(x,\, y,\, z) \tag{88}$$

requires the use of formula (82_2) of § 17 [173] with $n = 4$, ω being replaced by $\omega(x_2, x_3, x_4)\, e^{cx_1}$. After some simple rearrangements, we

obtain the solution of (87) with initial conditions (88) in the form:

$$u = \frac{1}{t}\frac{\partial}{\partial t}\int_0^t \varrho^2 I_0\left(ic\sqrt{t^2-\varrho^2}\right) T_\varrho\left\{\omega(x, y, z)\right\} d\varrho ,$$

where $T_\varrho\{\omega(x, y, z)\}$ is as usual the mean of $\omega(x, y, z)$ over the sphere of radius ϱ and centre at (x, y, z).

§ 19. The vibrations of rods

189. Fundamental equations. Fourier's method is used for a number of problems of mathematical physics which lead to partial differential equations. We arrive at expansions of a given function into the functions originating from application of the method. We have had examples of such expansions in the series of problems worked out above.

A further example is presented by the transverse vibrations of a rod, the equations for which we shall now deduce.

A thin rod differs from a string in that it does work on bending. The required function is the ordinate $y(x, t)$ of the axis of deformation of the rod with abscissa x and at the instant t.

If M is the bending moment and $F(x, t)$ the loading per unit length, we know [16] from the theory of flexure that

$$EI\frac{\partial^2 y}{\partial x^2} = M; \quad \frac{\partial^2 M}{\partial x^2} = F, \tag{1}$$

so that we find on differentiating the first equation twice with respect to x:

$$EI\frac{\partial^4 y}{\partial x^4} = F(x, t). \tag{2}$$

Equation (2) would express the condition of *equilibrium* of the rod if the force F were independent of time and the rod remained at rest. To obtain the equation of *motion*, we have to include the inertia force per unit length along with the external force, in accordance with d'Alembert's principle. The acceleration of a section x can be reckoned constant at all points of the section and equal to $\partial^2 y/\partial t^2$, the inertia force being evidently found by multiplying the acceleration by ϱS, where ϱ is the *volumetric* density of the material of the rod and S is the cross-sectional area of the section. We must thus replace F by $F - \varrho S\, \partial^2 y/\partial t^2$ in equation (2), which gives us the fourth order equation:

$$\frac{\partial^2 y}{\partial t^2} + b^2\frac{\partial^4 y}{\partial x^4} = f(x, t), \tag{3}$$

where

$$b^2 = \frac{EI}{\varrho S}, \; f(x, t) = \frac{1}{\varrho S} F(x, t). \tag{4}$$

Great importance is attached to the *boundary conditions* which must be satisfied at the ends $x = 0$ and $x = l$ of the rod, the form of these being dependent on how the respective ends are fixed. If an end is *fixed rigidly*, so that the rod has a horizontal direction at that end, we get the two conditions:

$$y = 0, \; \frac{\partial y}{\partial x} = 0 \quad \text{for} \quad x = 0 \quad \text{or} \quad x = l. \tag{5}$$

If an end is merely *supported*, i.e. free turning is possible about the fixing point, the bending moment must be zero at this point, i.e. we have the conditions

$$y = 0, \; \frac{\partial^2 y}{\partial x^2} = 0 \quad \text{for} \quad x = 0 \quad \text{or} \quad x = l. \tag{6}$$

Finally, if an end is *free*, both the bending moment and the shearing force $\partial M/\partial x$ must be zero at that end, whereas y itself may now differ from zero. Hence in this case

$$\frac{\partial^2 y}{\partial x^2} = 0, \; \frac{\partial^3 y}{\partial x^3} = 0 \quad \text{for} \quad x = 0 \quad \text{or} \quad x = l. \tag{7}$$

In all these cases, we get *a pair of conditions for each end*, contrary to the string, where we had one condition per end.

We also have to bring in *initial conditions*, of the same type as we had with a string:

$$y\Big|_{t=0} = \varphi(x), \quad \frac{\partial y}{\partial t}\Big|_{t=0} = \varphi_1(x). \tag{8}$$

As regards the free vibrations, we put $f(x, t) = 0$ in equation (3), which gives

$$\frac{\partial^2 y}{\partial t^2} + b^2 \frac{\partial^4 y}{\partial x^4} = 0. \tag{9}$$

190. Particular solutions. As in the case of a string, we seek a particular solution of this equation in the form

$$y = T(t) X(x). \tag{10}$$

We find by substituting in (9) that

$$T''(t) X(x) + b^2 T(t) X^{(4)}(x) = 0,$$

or, as for a string:

$$\frac{T''(t)}{b^2T(t)} = -\frac{X^{(4)}(x)}{X(x)} = -k^4,$$

where k^4 is constant, k being assumed real.

This gives us:

$$T''(t) + b^2k^4 T(t) = 0, \tag{11}$$

$$X^{(4)}(x) - k^4 X(x) = 0. \tag{12}$$

The general solution of equation (11) is

$$T(t) = N \sin(bk^2t + \varphi), \tag{13}$$

i.e. solution (10) is again a standing wave, such that points of the rod perform harmonic vibrations of the same frequency and phase but of different amplitudes $NX(x)$ which depend on x.

The general solution of equation (12) can also readily be found. Its characteristic equation [30] is

$$a^4 - k^4 = 0$$

with roots k, $-k$, ik, $-ik$ for $k \neq 0$, so that its general solution is

$$X(x) = C_1 e^{kx} + C_2 e^{-kx} + C_3 \cos kx + C_4 \sin kx, \tag{14}$$

which may be written alternatively, on expressing e^{kx} and e^{-kx} in terms of $\cosh kx$ and $\sinh kx$ and changing the arbitrary constants C_1 and C_2:

$$X(x) = C_1 \cosh kx + C_2 \sinh kx + C_3 \cos kx + C_4 \sin kx. \tag{15}$$

We now distinguish the various cases of boundary conditions.

1. If the rod is supported at both ends, we must satisfy conditions (6) at $x = 0$ and $x = l$, i.e.

$$X(0) = C_1 + C_3 = 0; \quad X''(0) = k^2(C_1 - C_3) = 0$$

$$X(l) = C_1 \cosh kl + C_2 \sinh kl + C_3 \cos kl + C_4 \sin kl = 0$$

$$X''(l) = k^2(C_1 \cosh kl + C_2 \sinh kl - C_3 \cos kl - C_4 \sin kl) = 0,$$

and these obviously give with $k \neq 0$:

$$C_1 = C_3 = 0; \tag{16}$$

$$C_2 \sinh kl + C_4 \sin kl = 0; \quad C_2 \sinh kl - C_4 \sin kl = 0. \tag{16$_1$}$$

The last system of simultaneous equations has the obvious solution $C_2 = C_4 = 0$, in which case all the constants C are zero and we get the trivial solution $X(x) = 0$. We neglect this case and assume that at least one of constants C_2, C_4 differs from zero.

If $C_4 = 0$, it follows from (16_1) that $C_2 = 0$ since $\sinh kl \neq 0$ for $k \neq 0$ [I, 177]. We must therefore take $C_4 \neq 0$. We get $C_4 \sin kl = 0$ on subtracting the two equations (16_1), so that we finally have the equation for k:

$$\sin kl = 0.$$

If this condition is satisfied, equations (16_1) now reduce to $C_2 \sinh kl = 0$, so that $C_2 = 0$; thus we obtain from (14), on writing $C_4 = C$:

$$X(x) = C \sin kx. \tag{17}$$

Equation (17) gives the same values for k:

$$\frac{\pi}{l}, \frac{2\pi}{l}, \ldots, \frac{n\pi}{l}, \ldots,$$

as in the case of a string, and the subsequent arguments and formulae are the same as in [167], the only difference being that the frequency ω_n of the nth harmonic is given by

$$\omega_n = \frac{bn^2\pi^2}{l^2}, \tag{18}$$

instead of by expression (44) [168].

With $k = 0$, equation (12) has the general solution: $X(x) = C_1 + C_2 x + C_3 x^2 + C_4 x^3$, and we find on attempting to satisfy conditions (6) that all the constants C must vanish.

2. If the rod is rigidly constrained at both ends, we must satisfy conditions (5) for $x = 0$ and $x = l$, which gives:

$$X(0) = C_1 + C_3 = 0; \quad X'(0) = k(C_2 + C_4) = 0$$
$$X(l) = C_1 \cosh kl + C_2 \sinh kl + C_3 \cos kl + C_4 \sin kl = 0$$
$$X'(l) = k(C_1 \sinh kl + C_2 \cosh kl - C_3 \sin kl + C_4 \cos kl) = 0.$$

Hence it is clear that

$$C_3 = -C_1; \quad C_4 = -C_2, \tag{19}$$

and we get the system of simultaneous equations for C_1 and C_2:

$$\left.\begin{aligned} C_1(\cosh kl - \cos kl) + C_2(\sinh kl - \sin kl) = 0 \\ C_1(\sinh kl + \sin kl) + C_2(\cosh kl - \cos kl) = 0. \end{aligned}\right\} \tag{20}$$

The necessary and sufficient condition for this system to have a solution differing from $C_1 = C_2 = 0$ is that the coefficients of C_1, C_2 are proportional:

$$\frac{\cosh kl - \cos kl}{\sinh kl + \sin kl} = \frac{\sinh kl - \sin kl}{\cosh kl - \cos kl}.$$

In this case, the two equations (20) reduce to one, which can be written, on using the relationships:

$$\cos^2 x + \sin^2 x = 1\,, \quad \cosh^2 x - \sinh^2 x = 1\,,$$

in the form:

$$\cosh kl \cdot \cos kl = 1\,. \tag{21}$$

We have obtained an equation for k analogous to (17) in the previous case. On writing

$$kl = \lambda\,,$$

for brevity, we have a transcendental equation for λ:

$$\cosh \lambda \cos \lambda = 1\,. \tag{22}$$

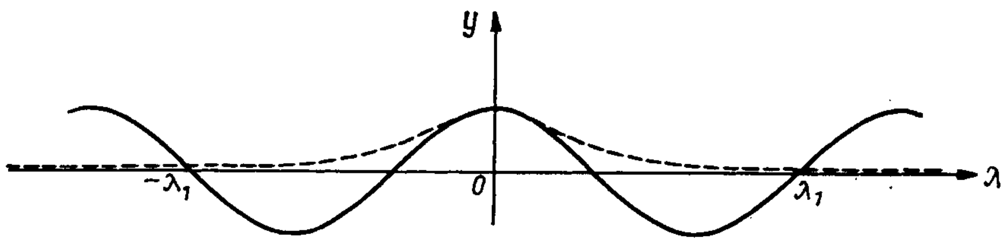

FIG. 140

On writing (22) as

$$\cos \lambda = \frac{1}{\cosh \lambda}$$

and drawing the graphs of the left and right-hand sides (Fig. 140), we find that (22) has an infinite set of real roots:

$$0\,,\ \pm\lambda_1,\ \pm\lambda_2, \ldots,\ \pm\lambda_n, \ldots$$

where the difference

$$\lambda_n - \frac{2n+1}{2}\pi \to 0 \quad \text{as} \quad x \to \infty\,.$$

We shall only pay attention to the positive roots for the present:

$$\lambda_1,\ \lambda_2, \ldots,\ \lambda_n, \ldots \tag{23}$$

These correspond to an infinity of values of parameter k:

$$k_1, k_2, \ldots k_n, \ldots;\ k_n = \frac{\lambda_n}{l} \ldots \tag{24}$$

Condition (21) is satisfied for these values of k, one of equations (20) is implied by the other, and we can put:

$$C_1 = C(\sinh kl - \sin kl);\ C_2 = -C(\cosh kl - \cos kl).$$

On writing for constants C_3, C_4 their values given by (19), substituting in (15) and setting $C = 1$, which we can obviously can do without loss of generality, we obtain the solution $X(x)$ as:

$$X(x) = (\sinh kl - \sin kl)(\cosh kx - \cos kx) \\ - (\cosh kl - \cos kl)(\sinh kx - \sin kx). \tag{25}$$

Strictly speaking, we get the infinite set of solutions,

$$X_1(x),\ X_2(x),\ \ldots,\ X_n(x),\ \ldots, \tag{26}$$

found by substituting k_n for k in general formula (25).

We cannot make use of the negative roots, $-\lambda_1, -\lambda_2, \ldots$ since they correspond with the values $-k_1, -k_2, \ldots$ of the parameter, which yield the same sequence of functions (26) in view of the oddness of functions (25) with respect to k.

On replacing k by its values (24) in (13), we find the corresponding sequence of functions $T(t)$:

$$T_1(t), T_2(t), \ldots, T_n(t), \ldots; T_n(t) = N_n \sin(\omega_n t + \varphi_n);\ \omega_n = bk_n^2, \tag{27}$$

and finally, the sequence of solutions of equation (9):

$$y_1(x,t), y_2(x,t), \ldots, y_n(x,t), \ldots; y_n(x,t) = T_n(t)\, X_n(x). \tag{28}$$

We obtain exactly analogous results for all the remaining conditions for the ends of the rod: having expressed the function $X(x)$ in the form (15) and substituted it in the boundary conditions, we get a system of four simultaneous equations with four unknowns C_1, C_2, C_3, C_4, which have a non-zero solution when and only when the parameter k satisfies a certain transcendental equation having an infinity of real roots. Substitution of a root k of this equation in the coefficients of the system gives a system in which one equation is a consequence of the other three, and constants C_1, C_2, C_3, C_4 are completely determined except for an arbitrary common factor: thus we obtain the functions $X_n(x)$ as linear combinations of ordinary and hyperbolic sines and cosines.

191. The expansion of an arbitrary function. We shall not stop to work out all the particular cases in detail but shall consider how to satisfy the *initial conditions:*

$$y\Big|_{t=0} = \varphi(x), \quad \frac{\partial y}{\partial t}\Big|_{t=0} = \varphi_1(x). \tag{29}$$

We use the same method as in the case of a string and seek $y(x, t)$ as a sum of particular solutions (28):

$$y(x,t) = \sum_{n=1}^{\infty} y_n(x,t) = \sum_{n=1}^{\infty} T_n(t)\, X_n(x). \tag{30}$$

On setting:

$$a_n = N_n \sin\varphi_n,\ \ b_n = N_n \cos\varphi_n,$$

(30) becomes:

$$y(x,t) = \sum_{n=1}^{\infty} (a_n \cos\omega_n t + b_n \sin\omega_n t)\, X_n(x), \tag{30$_1$}$$

and conditions (29) give:

$$\sum_{n=1}^{\infty} a_n X_n(x) = \varphi(x); \ \ \sum_{n=1}^{\infty} b_n \omega_n X_n(x) = \varphi_1(x). \tag{31}$$

We see from this that the problem of finding the coefficients a_n, b_n reduces to that of expanding the given functions $\varphi(x)$, $\varphi_1(x)$ into series in the functions $X_n(x)$. These latter series are analogous to the Fourier series considered above.

We shall follow the same method as for the Fourier series in [142] and merely show how to find the coefficients of these expansions; we assume that the expansions are possible and omit any discussion of their convergence or divergence. We also assume here that the boundary conditions of the problem are not necessarily those enumerated in items 1 and 2 of [190], but are any of (5), (6) and (7) above.

Let $f(x)$ be a function given in the interval $(0, l)$ and let it be required to expand this in the form:

$$f(x) = \sum_{n=1}^{\infty} A_n X_n(x). \tag{32}$$

We assume that this expansion is possible and that series (32) can be integrated term by term. The coefficients A_n can be found by virtue of the orthogonality of the functions

$$X_1(x),\ X_2(x),\ \ldots,\ X_n(x),\ \ldots$$

in the interval $(0, l)$; for we shall now show that

$$\int_0^l X_n(x)\, X_m(x)\, \mathrm{d}x = 0, \quad \text{if} \quad n \neq m. \tag{33}$$

We do this by first noting that the $X_n(x)$ satisfy equation (12) by definition, if the k there is replaced by k_n, i.e.

$$X_n^{(4)}(x) = k_n^4 X_n(x).$$

We now have:

$$X_n^{(4)}(x) = k_n^4 X_n(x); \;\; X_m^{(4)}(x) = k_m^4 X_m(x). \tag{34}$$

On multiplying the first equation by $X_m(x)$ and the second by $X_n(x)$, then subtracting and integrating from 0 to l with respect to x, we get:

$$(k_m^4 - k_n^4) \int_0^l X_n(x)\, X_m(x)\, \mathrm{d}x =$$

$$= \int_0^l \left[X_m^{(4)}(x)\, X_n(x) - X_n^{(4)}(x)\, X_m(x)\right] \mathrm{d}x, \tag{35}$$

After this, we only have to show, in order to prove (33), that

$$\int_0^l \left[X_m^{(4)}(x)\, X_n(x) - X_n^{(4)}(x)\, X_m(x)\right] \mathrm{d}x = 0, \tag{36}$$

since the factor $(k_m^4 - k_n^4)$ does not vanish for $m \neq n$.

Integration by parts gives us:

$$\int X_m^{(4)}(x)\, X_n(x)\, \mathrm{d}x = X_m'''(x)\, X_n(x) - \int X_m'''(x)\, X_n'(x)\, \mathrm{d}x =$$
$$= X_m'''(x)\, X_n(x) - X_m''(x)\, X_n'(x) + \int X_m''(x)\, X_n''(x)\, \mathrm{d}x,$$

and similarly:

$$\int X_n^{(4)}(x)\, X_m(x)\, \mathrm{d}x = X_n'''(x)\, X_m(x) - X_n''\, X_m'(x) + \int X_n''(x)\, X_m''(x)\, \mathrm{d}x,$$

whence we readily deduce that

$$\int_0^l \left[(X_m^{(4)}(x)\, X_n(x) - X_n^{(4)}(x)\, X_m(x)\right] \mathrm{d}x =$$

$$= \left[X_m'''(x)\, X_n(x) - X'''(x)\, X_m(x)\right] \Big|_{x=0}^{x=l} -$$

$$- \left[X_m''(x)\, X_n'(x) - X_n''(x)\, X_m'(x)\right] \Big|_{x=0}^{x=l}.$$

The right-hand side of this last equation contains the values at $x = 0$ and $x = l$ of $X_m(x)$, $X_n(x)$ and their derivatives up to and including the third order, and whichever of conditions (5), (6) or (7)

we take, a vanishing factor is found in each term of the right-hand side. Thus equation (36) is proved, and with this the orthogonality expressed by (33).

With $m = n$, integral (33) becomes

$$I_n = \int_0^l X_n^2(x)\,dx \tag{37}$$

and is a well-defined constant which can readily be found in each particular case. We have, for instance, in case 1 [190]:

$$I_n = \int_0^l \sin^2 \frac{n\pi x}{l}\,dx = \frac{l}{2}.$$

It follows that if we change the system of functions $X_1(x)$, $X_2(x)$, $\ldots$, $X_n(x)$ into the system

$$\frac{X_1(x)}{\sqrt{I_1}},\ \frac{X_2(x)}{\sqrt{I_2}},\ \ldots,\ \frac{X_n(x)}{\sqrt{I_n}},\ \ldots,$$

we obtain a *normalized* as well as *orthogonal* system [148], i.e. the integral of the square of each function is unity. We return to the determination of the coefficients A_n of expansion (32). On multiplying both sides by $X_m(x)$, integrating with respect to x from 0 to l, and taking account of relationships (33) and (37), we find at once that

$$\int_0^l f(x)\,X_m(x)\,dx = A_m I_m,$$

whence

$$A_m = \frac{\int_0^l f(x)\,X_m(x)\,dx}{I_m} = \frac{\int_0^l f(x)\,X_m(x)\,dx}{\int_0^l X_m^2(x)\,dx}.$$

We thus arrive at an expansion resembling a Fourier series, of an arbitrary function $f(x)$:

$$f(x) = \sum_{n=1}^{\infty} A_n X_n(x), \quad \text{where} \quad A_n = \frac{\int_0^l f(x)\,X_n(x)\,dx}{\int_0^l X_n^2(x)\,dx}. \tag{38}$$

After what has been said, the determination of the constants a_n and b_n in equations (31) presents no difficulty at all; we have, in fact, on replacing $f(x)$ by $\varphi(x)$ and $\varphi_1(x)$ in (38):

$$a_n = \frac{\int\limits_0^l \varphi(x) X_n(x)\,dx}{\int\limits_0^l X_n^2(x)\,dx}, \quad b_n = \frac{\int\limits_0^l \varphi_1(x) X_n(x)\,dx}{\omega_n \int\limits_0^l X_n^2(x)\,dx}. \tag{39}$$

Substitution of all these in series (30) gives us the final solution of the problem.

The forced vibrations of a rod are treated precisely as in the case of a string, except that the function $f(x, t)$ is now expanded in functions $X_n(x)$ instead of in sines.

It is clear from the above that the standing wave method is applicable with equal success to the vibrations both of strings and of rods. Whilst the method of characteristics is very useful in studying the vibrations of strings and the equation of telegraphy, it has not yet been applied with real success to equation (9).

§ 20. Laplace's equation

192. Harmonic functions. We consider in the present article partial differential equations of the form

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0, \tag{1}$$

where U is a function of x, y, z. As already mentioned, (1) is known as Laplace's equation. We have also seen above that the left-hand side is written symbolically as ΔU, this being described as Laplace's operator on the function U. We saw in [87] that equation (1) must be satisfied by the potential of a gravitational force or of the interaction between electric charges at all points of space outside the attracting bodies or the charges which produce the field.

An equation of type (1) was also encountered in [114], where it was satisfied by the velocity potential of the irrotational flow of an incompressible fluid. We proved in [117] that (1) must likewise be satisfied by the temperature in a homogeneous body if the heat exchange is stationary, i.e. the temperature U depends only on the position of the point and not on time. Similarly, our investigation of [118] of a stationary electromagnetic field led to Laplace's equation.

If U is independent of one coordinate, say z, equation (1) reduces to

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0. \tag{2}$$

In this case, U has the same value on any line parallel to the z axis, in other words, the values of U show the same variational picture on any plane parallel to the xy plane, so that only the latter plane need be considered.

A function which is continuous, together with its derivatives up to the second order, in a volume (three-dimensional domain) (D) where it satisfies equation (1) is said to be *harmonic in* (D). The same term is applied in regard to equation (2) for a domain in the xy plane. We elucidate below some properties of harmonic functions.

A function U, in addition to satisfying equation (1), has usually to be subjected to certain boundary conditions in problems of mathematical physics. Initial conditions are naturally absent in the present case. Boundary value problems for equation (1) amount fundamentally to the following: to find a function which is harmonic in a domain (D) and the values of which are assigned on a surface (S) of the domain. This is generally known as *Dirichlet's problem.* When we speak of the values of U "on a surface (S) of the domain" we understand the limiting values attained by U on approaching points of (S) from inside (D). The problem may be formulated more precisely as: to seek a function U which is harmonic inside (D) and is continuous over (D) including its boundary (S), the values of U being specified on (S). The function specified on (S) must naturally be continuous. We shall assume for simplicity that the boundary of (D) is a single closed surface (S). It may be noted that (D) can be finite or infinite. In the latter case, it lies outside (S). With a finite domain, we have an *interior Dirichlet problem,* and with an infinite domain, an *exterior Dirichlet problem.* The latter problem requires the further condition that the function tends to zero at an infinite distance, or as we usually say, vanishes at infinity. The boundary condition for Dirichlet's problem is written as

$$U\,|_{(S)} = f(M), \tag{3}$$

where $f(M)$ is a continuous function given on the surface (S) and M is a variable point of (S). The interior Dirichlet problem is similarly stated in regard to equation (2) for a plane domain, the boundary condition consisting of a specification of U on the contour of the domain. The exterior Dirichlet problem on a plane requires the function to have a finite limit at an infinite distance.

We shall mention one other type of boundary condition, when the normal derivative is assigned on the surface (S):

$$\left.\frac{\partial U}{\partial n}\right|_{(S)} = f(M). \tag{4}$$

The task of finding a harmonic function satisfying this type of boundary condition is known as *Neumann's problem*. It is met with in hydrodynamics when considering the motion of a rigid body in an ideal incompressible fluid. Boundary condition (4) here expresses the equality of the normal components of velocity of a point M of surface (S) of the body and of the fluid particle adjacent to M. Neumann's problem can likewise be stated for equation (2).

We shall derive some necessary formulae before passing on to elucidate the properties of harmonic functions.

193. Green's formula. Let (D) be a finite domain bounded by a surface (S), and let U, V be two functions which are continuous and have continuous derivatives up to the second order inside (D) up to its boundary (S). We consider the integral:

$$I = \iiint\limits_{(D)} \left(\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z}\right) dv = \iiint\limits_{(D)} \operatorname{grad} U \cdot \operatorname{grad} V \, dv.$$

On using the obvious identity:

$$\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} = \frac{\partial}{\partial x}\left(U \frac{\partial V}{\partial x}\right) - U \frac{\partial^2 V}{\partial x^2}$$

and the corresponding identities for $\partial/\partial y$ and $\partial/\partial z$, we can write the integral as

$$I = \iiint\limits_{(D)} \left[\frac{\partial}{\partial x}\left(U \frac{\partial V}{\partial x}\right) + \frac{\partial}{\partial y}\left(U \frac{\partial V}{\partial y}\right) + \frac{\partial}{\partial z}\left(U \frac{\partial V}{\partial z}\right)\right] dv - \iiint\limits_{(D)} U \,\Delta V \, dv.$$

We transform the first term on the right by using Ostrogradskii's formula:

$$I = \iint\limits_{(S)} \left[U \frac{\partial V}{\partial x} \cos(n, X) + U \frac{\partial V}{\partial y} \cos(n, Y) + \right.$$

$$\left. + U \frac{\partial V}{\partial z} \cos(n, Z)\right] dS - \iiint\limits_{(D)} U \,\Delta V \, dv$$

or [102]:

$$I = \iint\limits_{(S)} U \frac{\partial V}{\partial n} dS - \iiint\limits_{(D)} U \,\Delta V \, dv,$$

where (n) is the outward normal in respect to (D) at points of the surface (S).

We thus arrive at what is known as Green's preliminary formula:

$$\iiint\limits_{(D)} \left(\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z}\right) \mathrm{d}v =$$

$$= \iiint\limits_{(D)} \operatorname{grad} U \cdot \operatorname{grad} V \, \mathrm{d}v = \iint\limits_{(S)} U \frac{\partial V}{\partial n} \mathrm{d}S - \iiint\limits_{(D)} U \Delta V \, \mathrm{d}v. \tag{5}$$

The left-hand side of this equation remains the same on interchanging U and V, so that the same must be true for the right-hand side, i.e. we can write:

$$\iint\limits_{(S)} U \frac{\partial V}{\partial n} \mathrm{d}S - \iiint\limits_{(D)} U \Delta V \, \mathrm{d}v = \iint\limits_{(S)} V \frac{\partial U}{\partial n} \mathrm{d}S - \iiint\limits_{(D)} V \Delta U \, \mathrm{d}v,$$

whence Green's formula is obtained in its final form:

$$\iiint\limits_{(D)} (U \Delta V - V \Delta U) \, \mathrm{d}v = \iint\limits_{(S)} \left(U \frac{\partial V}{\partial n} - V \frac{\partial U}{\partial n}\right) \mathrm{d}S. \tag{6}$$

We sometimes use the inward instead of the outward normal, in which case we only have to change the signs of the derivatives with respect to the normal on the right-hand side; thus Green's formula reads with the inward normal:

$$\iiint\limits_{(D)} (U \Delta V - V \Delta U) \, \mathrm{d}v = \iint\limits_{(S)} \left(V \frac{\partial U}{\partial n_i} - U \frac{\partial V}{\partial n_i}\right) \mathrm{d}S, \tag{6_1}$$

where n_i is the normal direction into (D).

The domain (D) can be bounded by a number of surfaces (S). Green's formula is applicable in this case except that the surface integral on the right has to be taken over all the surfaces bounding (D). It should be noted that the outward normal (n) from (D) is now directed into the surfaces that bound (D) from the inside [63].

As we have remarked, it is sufficient to require the continuity of functions U, V and of their derivatives up to the second order as far as the surface (S) when deducing Green's formula (6). Certain demands must naturally be imposed on (S). We may fall back here on the conditions for which Ostrogradskii's formula was deduced [63]. These conditions amounted to the following: the surface (S) can be split up into a finite number of pieces such that there is a continuously

varying tangent plane on each piece as far as its boundary. Such a surface is said to be piecewise smooth. The boundaries of the pieces of surface in question must be piecewise smooth lines. This condition imposed on the surface can likewise be expressed analytically.

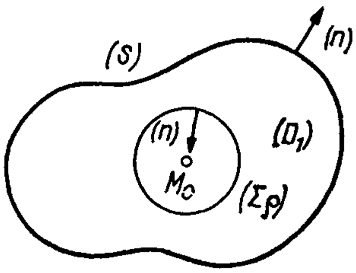

FIG. 141

An important practical corollary of Green's formula expresses the value of the function at any given point M_0 inside (D) as the sum of a surface and a volume integral. Let the function $U(M)$ be defined in the domain (D) and be continuous along with its derivatives up to the second order as far as (S).

We apply Green's formula to this function and to the function $V = 1/r$, where r is the distance from a given point M_0 inside (D) to a variable point M. The function $V = 1/r$ becomes infinite if the point M coincides with M_0, and we cannot apply Green's formula to the whole of (D). We isolate M_0 with a small sphere of centre M_0 and small radius ϱ, and write (Σ_ϱ) for the surface of this sphere and (D_1) for what is left of (D) after removing the sphere (Fig. 141). The functions U and $V = 1/r$ have the required continuity in (D_1) and we can apply Green's formula to this domain, which gives us:

$$\iiint\limits_{(D_1)} \left[U\Delta\left(\frac{1}{r}\right) - \frac{1}{r}\Delta U\right] \mathrm{d}v =$$

$$= \iint\limits_{(S)} \left[U\frac{\partial\left(\frac{1}{r}\right)}{\partial n} - \frac{1}{r}\frac{\partial U}{\partial n}\right] \mathrm{d}S + \iint\limits_{(\Sigma_\varrho)} \left[U\frac{\partial\left(\frac{1}{r}\right)}{\partial n} - \frac{1}{r}\frac{\partial U}{\partial n}\right] \mathrm{d}S, \quad (7)$$

the integration being carried out over both the surfaces (S) and (Σ_ϱ) bounding (D_1). But, as we have seen, $V = 1/r$ satisfies Laplace's equation, i.e. $\Delta(1/r) = 0$ [119]. Furthermore, the normal n is in precisely the opposite direction to the radius r on the sphere (Σ_ϱ), so that the normal derivative in the integration over (Σ_ϱ) is taken with the opposite sign to the derivative with respect to r. On taking all this into account, we can write (7) in the form:

$$\iiint\limits_{(D_1)} \frac{\Delta U}{r} \mathrm{d}v + \iint\limits_{(S)} \left[U\frac{\partial\left(\frac{1}{r}\right)}{\partial n} - \frac{1}{r}\frac{\partial U}{\partial n}\right] \mathrm{d}S +$$

$$+ \iint\limits_{(\Sigma_\varrho)} \frac{1}{r^2} U\,\mathrm{d}S - \iint\limits_{(\Sigma_\varrho)} \frac{1}{r}\frac{\partial U}{\partial n}\mathrm{d}S = 0. \quad (8)$$

We now let the radius ϱ of the sphere tend to zero. In this case, the first term in the last equation tends to the volume integral over the whole of (D) [86]. The second term is independent of ϱ. We show that the third term tends to the limit $4\pi\, U(M_0)$. Using the fact that r has the constant value ϱ on (Σ_ϱ), we can write:

$$\iint\limits_{(\Sigma_\varrho)} \frac{1}{r^2} U(M)\, \mathrm{d}S = \frac{1}{\varrho^2} \iint\limits_{(\Sigma_\varrho)} U(M)\, \mathrm{d}S.$$

The mean value theorem gives us:

$$\iint\limits_{(\Sigma_\varrho)} \frac{1}{r^2} U(M)\, \mathrm{d}S = \frac{1}{\varrho^2} U(M_\varrho) \cdot 4\pi\varrho^2 = 4\pi U(M_\varrho),$$

where M_ϱ is a point on the surface of (Σ_ϱ). This point tends to M_0 as $\varrho \to 0$, whence it is clear that the above expression tends to $4\pi\, U(M_0)$. Similarly, application of the mean value theorem to the last term gives us:

$$-\iint\limits_{(\Sigma_\varrho)} \frac{1}{r} \frac{\partial U}{\partial n}\, \mathrm{d}S = -\frac{1}{\varrho} \iint\limits_{(\Sigma_\varrho)} \frac{\partial U}{\partial n}\, \mathrm{d}S = -\frac{1}{\varrho} \frac{\partial U}{\partial n}\bigg|_{M_\varrho} 4\pi\varrho^2 = -\frac{\partial U}{\partial n}\bigg|_{M_\varrho} 4\pi\varrho.$$

The first order derivative of U with respect to any direction remains bounded as M_ϱ tends to M_0, since by hypothesis U has continuous derivatives up to the second order everywhere inside (D). The factor $4\pi\varrho$ tends to zero as $\varrho \to 0$. Hence it is clear that the last term in equation (8) tends to zero. The limit of equation (8) finally gives the required corollary of Green's formula:

$$\iiint\limits_{(D)} \frac{\mathrm{d}U}{r} \Delta v + \iint\limits_{(S)} \left[U \frac{\partial \left(\frac{1}{r}\right)}{\partial n} - \frac{1}{r} \frac{\partial U}{\partial n} \right] \mathrm{d}S + 4\pi U(M_0) = 0$$

or

$$U(M_0) = \frac{1}{4\pi} \iint\limits_{(S)} \left[\frac{1}{r} \frac{\partial U}{\partial n} - U \frac{\partial \left(\frac{1}{r}\right)}{\partial n} \right] \mathrm{d}S - \frac{1}{4\pi} \iiint\limits_{(D)} \frac{\Delta U}{r}\, \mathrm{d}v. \tag{9}$$

We notice once more that this formula is valid for any function U which is continuous and has continuous derivatives to the second order in the domain (D) as far as (S).

Similar formulae are applicable in the case of a plane. We shall state these without dwelling on their proof. Let (B) be a plane domain with contour (l), and let n be the outward normal with respect to (B)

to the contour. In the case of a plane, Laplace's operator takes the form in Cartesian coordinates:

$$\Delta U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2}.$$

The formula for a plane corresponding to (6) runs:

$$\iint\limits_{(B)} (U\Delta V - V\Delta U)\,\mathrm{d}S = \int\limits_{(l)} \left(U\frac{\partial V}{\partial n} - V\frac{\partial U}{\partial n}\right)\mathrm{d}s. \tag{10}$$

There is no complete analogy in regard to expression (9) since it is essential for (9) that the function $1/r$ satisfies Laplace's equation. This is not true on a plane, and instead of $1/r$ we have to take a solution of Laplace's equation of the form $\log r$ or $\log(1/r) = -\log r$, where r is the distance from any fixed point of the plane to a variable point M. Thus instead of (9), we have the formula on a plane:

$$U(M_0) = \frac{1}{2\pi}\int\limits_{(l)} \left[U\frac{\partial(\log r)}{\partial n} - \log r\frac{\partial U}{\partial n}\right]\mathrm{d}s + \frac{1}{2\pi}\iint\limits_{(B)} \Delta U\cdot\log r\,\mathrm{d}S, \tag{11}$$

where M_0 is any fixed point inside (B) and r is the distance of a variable point M from M_0.

It may be noticed that the triple integral in expression (9) is improper, since the integrand becomes infinite at the point M_0. The integral is obviously convergent, however, since the absolute value of the integrand is less that A/r^p with $p = 1$. A similar remark applies in the case of expression (11).

194. The fundamental properties of harmonic functions. We take a function U, harmonic in a bounded domain (D) with surface (S). On assuming that U and its derivatives to the second order are continuous as far as (S), we can apply Green's formula (6) to U and the harmonic function $V = 1$, and obtain, since $\Delta V = \Delta(1) = 0$ and $\partial(1)/\partial n = 0$:

$$\iint\limits_{(S)} \frac{\partial U}{\partial n}\,\mathrm{d}S = 0, \tag{12}$$

which gives us the first property of harmonic functions: *the integral of the normal derivative of a harmonic function over the surface of the domain is zero.*

If we apply expression (9) to the harmonic function U, we obtain, since $\Delta U = 0$:

$$U(M_0) = \frac{1}{4\pi}\iint\limits_{(S)} \left[\frac{1}{r}\frac{\partial U}{\partial n} - U\frac{\partial\left(\frac{1}{r}\right)}{\partial n}\right] \mathrm{d}s. \tag{13}$$

Hence we have a second property of harmonic functions: *the value of a harmonic function at any interior point of the domain is expressed by equation* (13) *in terms of the values of the function and its normal derivative on the surface of the domain.*

We remark that the integrals in (12) and (13) do not contain second order derivatives of U. In order to apply these formulae, it is sufficient to assume continuity of the harmonic function and its first order derivate as far as (S). This may be seen simply by applying a slight contraction to the surface (S) and writing (12) and (13) for the contracted domain (D') in which there is continuity of the second order derivatives as far as the boundary surface; we then pass to the limit on expanding (D') back to (D). The contraction can be accomplished, say, by adding the same small length δ to the inward normal at each point of (S). The ends of the added lengths form the new (contracted) surface. The surface (S) must be such that, for all sufficiently small δ, the operation described leads to a surface which does not cut itself and which is piecewise smooth [193]. This matter is treated in more detail in Volume IV.

We shall apply (13) to a particular type of domain — a sphere of radius R and centre M_0; we assume, of course, that the function U is harmonic in the sphere and that both U and its first order derivatives are continuous as far as its surface (Σ_R).

In this case, the outward normal n has the same direction as the radius of the sphere, so that we have:

$$\frac{\partial\left(\frac{1}{r}\right)}{\partial n} = \frac{\partial\left(\frac{1}{r}\right)}{\partial r} = -\frac{1}{r^2},$$

and equation (13) gives:

$$U(M_0) = \frac{1}{4\pi}\iint\limits_{(\Sigma_R)} \left(\frac{1}{r}\frac{\partial U}{\partial n} + \frac{1}{r^2}U\right) \mathrm{d}S.$$

But r has the constant value R on the surface Σ_R of the sphere, so that:

$$U(M_0) = \frac{1}{4\pi R}\iint\limits_{(\Sigma_R)} \frac{\partial U}{\partial n}\,\mathrm{d}S + \frac{1}{4\pi R^2}\iint\limits_{(\Sigma_R)} U\,\mathrm{d}s,$$

or, by (12), we can finally write

$$U(M_0) = \frac{\iint\limits_{(\Sigma_R)} U \, \mathrm{d}s}{4\pi R^2}. \tag{14}$$

This formula expresses a third property of harmonic functions: *the value of a harmonic function at the centre of a sphere is equal to the arithmetic mean of its values on the surface of the sphere,* i.e. *is equal to its integral over the surface divided by the surface area.*

This leads us almost at once to a fourth property of harmonic functions:

A function which is continuous as far as the boundary of a domain and is harmonic in the interior attains its greatest and least value only on the boundary, unless the function is a constant. We give a detailed proof of this proposition. Let $U(M)$ attain its greatest value at some interior point M_1 of the domain D where $U(M)$ is harmonic. We draw a sphere Σ_ϱ with radius ϱ and centre M_1 so that it belongs to D, then we apply (14) and change the integrand U to its greatest value $U_\varrho^{(\max)}$ on Σ_ϱ. We thus get

$$U(M_1) \leqslant U_\varrho^{(\max)},$$

the sign of equality only being obtained in the case when U is constant on Σ_ϱ and equal to $U(M_1)$. Inasmuch as $U(M_1)$ is the greatest value of $U(M)$ in D by hypothesis, we must in fact obtain the sign of equality, so that we can say that $U(M)$ is constant inside and on the surface of every sphere with centre M_1 belonging to D. We show that it follows that $U(M)$ is constant throughout D.

Let N be any interior point of D. We have to show that $U(N) = U(M_1)$. We join M_1 and N by a line of finite length, say by a step-line, which lies inside D, and we let d be the shortest distance from this line l to the boundary S of D (d is a positive number). By what has been proved above, $U(M)$ is equal to the constant $U(M_1)$ on a sphere with centre M_1 and radius d. Let M_2 be the last point of intersection, reckoning from M_1, of the line l with the surface of this sphere. We have $U(M_2) = U(M_1)$, and by what was proved above, $U(M)$ is also equal to the constant $U(M_1)$ on the sphere with centre M_2 and radius d. Let M_3 be the last point of intersection of l with this sphere. As above, the function $U(M)$ is equal to the constant $U(M_1)$ on the sphere with centre M_3 and radius d, and so on. By constructing a finite number of such spheres, we can verify that $U(N) = U(M_1)$, which is what we had to prove. It can also be shown that

$U(M)$ can have neither maxima nor minima inside D. By using the demonstrated properties of harmonic functions, it is very easily shown that *the interior Dirichlet problem*, mentioned in [185], *can have only one solution*. In fact, if we suppose that two solutions $U_1(M)$ and $U_2(M)$ exist, harmonic inside D and having the same boundary values $f(M)$ on the boundary S of D, the difference $V(M) = U_1(M) - - U_2(M)$ will also satisfy Laplace's equation inside D, so that this is also a harmonic function, with a boundary value of zero everywhere on S. Hence it follows directly from what has been proved above that $V(M)$ is identically zero throughout D, since otherwise it would have to attain a positive greatest, value or negative least value inside D, which is impossible. Two solutions $U_1(M)$ and $U_2(M)$ of Dirichlet's problem must thus coincide throughout D. The uniqueness of the solution of the exterior Dirichlet problem is similarly proved, on taking into account the vanishing by hypothesis of harmonic functions at infinity.

Precisely analogous properties are obtained for harmonic functions on a plane. We have here, instead of (13):

$$U(M_0) = \frac{1}{2\pi} \int\limits_{l} \left(U \frac{\partial \log r}{\partial n} - \log r \frac{\partial U}{\partial n} \right) ds , \tag{15}$$

whilst the mean value theorem takes the form:

$$U(M_0) = \frac{1}{2\pi R} \int\limits_{\lambda_R} U \, ds , \tag{16}$$

where λ_R is the circle of radius R and centre M_0. For the exterior Dirichlet problem, we require only the existence of a finite limit at infinity, and not vanishing, as in the three-dimensional case, whilst the proof of uniqueness is different from that given above. The required proof will be found in Volume IV, where we consider the Dirichlet and Neumann problems in more detail.

We now notice that any constant is a harmonic function satisfying the boundary condition:

$$\left. \frac{\partial U}{\partial n} \right|_{\varrho} = 0 ,$$

whence it follows that the addition of an arbitrary constant does not affect the solution of Neumann's problem with given boundary values of $\partial U/\partial n$, i.e. the solution is defined to within an arbitrary constant. It also follows from equation (12) that the function $f(M)$

appearing in the boundary condition of the interior Dirichlet problem cannot be arbitrary, but must satisfy the condition:

$$\iint_S f(M)\,\mathrm{d}S = 0.$$

It may further be noted in conclusion that expression (13) still holds in the case when $U(M)$ is a harmonic function in an *infinite domain*, consisting of the part of space outside the surface S. Here, we only have to make a hypothesis about the order of smallness of $U(M)$ at infinity, i.e. when the point M becomes infinitely remote It is sufficient (and necessary) to assume the validity at infinity of the inequalities

$$R\,|U(M)| \leqslant A\,; \quad R^2\left|\frac{\partial U(M)}{\partial l}\right| \leqslant A\,, \tag{*}$$

where R is the distance of M from the origin or any other fixed point of space, A is a numerical constant, and l is an arbitrary direction in space. To prove relationship (13) for an infinite domain with the conditions indicated, we simply apply it for the finite domain bounded by the surface S and by a sphere of sufficiently large radius with centre at say M_0. As the radius tends to infinity, the integral over the spherical surface tends to zero in view of the conditions stated above, and we obtain (13) for any point M_0 lying outside S. As we shall see in Volume IV, conditions (*) are certainly satisfied if $U(M)$ tends to zero as M moves to infinity.

195. The solution of Dirichlet's problem for a circle. We saw in the previous section that the solution of Dirichlet's problem is unique. However, we do not yet know if a solution in general exists. We shall confine ourselves to particular cases and shall not consider existence in the general case. Various methods of solution will be used, and we start with the plane case.

Let it be required to find a function, harmonic inside a circle and taking previously assigned values on the circumference. Let R be the radius of the circle, and let the centre be taken as coordinate origin. The boundary values given on the circumference will now represent some known continuous function $f(\theta)$ of the polar angle on the circumference. We take a variable point M with polar coordinates (r, θ) inside the circle. The required function must satisfy Laplace's equation [119]:

$$\frac{\partial}{\partial r}\left(r\frac{\partial U}{\partial r}\right) + \frac{1}{r}\frac{\partial^2 U}{\partial \theta^2} = 0$$

or

$$r^2 \frac{\partial^2 U}{\partial r^2} + r \frac{\partial U}{\partial r} + \frac{\partial^2 U}{\partial \theta^2} = 0. \tag{17}$$

We apply Fourier's method here and seek the solution of (17) as the product of a function of θ only and a function of r only:

$$U = \chi(\theta) \cdot \omega(r). \tag{18}$$

We substitute this expression in (17):

$$r^2 \omega''(r)\, \chi(\theta) + r\, \omega'(r)\, \chi(\theta) + \chi''(\theta)\, \omega(r) = 0$$

or

$$\frac{\chi''(\theta)}{\chi(\theta)} = -\frac{r^2\omega''(r) + r\omega'(r)}{\omega(r)}. \tag{18_1}$$

The left-hand side of this last equation contains only the independent variable θ, and the right-hand side contains only r, so that both sides must be equal to the same constant which we can write as $(-k^2)$. We thus get two equations:

$$\chi''(\theta) + k^2\chi(\theta) = 0 \quad \text{and} \quad r^2\omega''(r) + r\omega'(r) - k^2\omega(r) = 0.$$

The first of these gives, for $k \neq 0$:

$$\chi(\theta) = A \cos k\theta + B \sin k\theta.$$

The second is Euler's equation [92]. We seek its solution as $\omega(r) = r^m$:

$$r^2 \cdot m(m-1)\, r^{m-2} + rmr^{m-1} - k^2 r^m = 0,$$

whence we obtain on cancelling $r^m : m^2 - k^2 = 0$, i.e. $m = \pm k$, so that the general solution becomes:

$$\omega(r) = Cr^k + Dr^{-k},$$

provided the constant k differs from zero. Substitution in expression (18) gives us for U:

$$U = (A \cos k\theta + B \sin k\theta)(Cr^k + Dr^{-k}). \tag{19}$$

With $k = 0$, we have the equations:

$$\chi''(\theta) = 0 \text{ and } r\omega''(r) + \omega'(r) = 0,$$

from which we readily find that

$$U = (A + B\theta)(C + D \log r). \tag{19_1}$$

We now proceed to determine the constants A, B, C, D and k appearing in expressions (19) and (19_1). We notice that the addition of 2π to the angle θ is equivalent to a rotation about the origin, in the course of which the single-valued function $U(r, \theta)$ must return to its initial value. Thus the first factor of (19), depending on θ, must be periodic in θ and of period 2π. It follows that the constant k can only take integral values, $k = \pm 1, \pm 2, \pm 3, \ldots, \pm n, \ldots$

But if we substitute $k = n$ or $k = -n$ in (19), the result obtained must be essentially the same in view of the arbitrariness of the coefficient B, so that we can confine ourselves to positive integral values of k (the characteristic numbers of the problem), i.e. $k = n$ $(n = 1, 2, 3, \ldots)$.

The periodicity of solution (19_1) requires the vanishing of the constant B. We thus arrive at the following solutions:

$$U_n(r, \theta) = (A_n \cos n\theta + B_n \sin n\theta)(C_n r^n + D_n r^{-n}) \quad (n = 1, 2, \ldots)$$

$$U_0(r, \theta) = A_0(C_0 + D_0 \log r),$$

where the constants have been provided with subscripts since they can be different for different values of n. Turning now to the second factor, that depends on r, we notice that the solution has to be finite and continuous at the centre of the circle, i.e. for $r = 0$. This implies that the constant D_0 and all the D_n have to be put equal to zero. On writing A_n for the arbitrary constants $A_n C_n$, and similarly B_n for $B_n C_n$ and $A_0/2$ for $A_0 C_0$, we can express the solutions as

$$U_n(r, \theta) = (A_n \cos n\theta + B_n \sin n\theta) r^n \quad (n = 1, 2, \ldots),$$

$$U_0(r, \theta) = \frac{A_0}{2}.$$

Since Laplace's equation is homogeneous and linear, the sum of these solutions will also be a solution, i.e. we arrive at the solution as:

$$U(r, 0) = \frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos \theta + B_n \sin n\theta) r^n. \tag{20}$$

We now determine the arbitrary constants A_n and B_n in accordance with the given boundary condition:

$$U(r, 0)\big|_{r=R} = f(\theta), \tag{21}$$

where $f(\theta)$ is a continuous function given in the interval $(-\pi \leqslant \theta \leqslant \pi)$ and $f(-\pi) = f(\pi)$.

This condition gives

$$f(\theta) = \frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos n\theta + B_n \sin n\theta) R^n. \tag{22}$$

It is clear from this that $A_n R^n$ and $B_n R^n$ are the Fourier coefficients of $f(\theta)$ in an interval of length 2π, say in the interval $(-\pi, \pi)$. On evaluating these from the familiar formulae

$$A_n = \frac{1}{\pi R^n} \int_{-\pi}^{+\pi} f(t) \cos nt \, dt; \;\; B_n = \frac{1}{\pi R^n} \int_{-\pi}^{+\pi} f(t) \sin nt \, dt \tag{23}$$

and substituting the values obtained in equation (2), we get the required solution of Dirichlet's problem.

On comparing Fourier series (22) with the solution as expressed by (20), we can state the result obtained as follows: *the solution of Dirichlet's problem for a circle is obtained by writing down the Fourier series for the boundary values $f(\theta)$ and multiplying the $(n+1)$-th term by $(r/R)^n$.*

The solution can be expressed as a definite integral instead of by the infinite series (20). We substitute expressions (23) for the coefficients in (20):

$$U(r, \theta) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) \, dt + \sum_{n=1}^{\infty} \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \cos n(t-\theta) \cdot \left(\frac{r}{R}\right)^n dt$$

or

$$U(r, \theta) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) \left[1 + 2\sum_{n=1}^{\infty} \left(\frac{r}{R}\right)^n \cos n(t-\theta)\right] dt.$$

Formula (14) of [I, 174] gives immediately:

$$1 + 2\sum_{n=1}^{\infty} r^n \cos n\varphi = \frac{1 - r^2}{r^2 - 2r\cos\varphi + 1} \quad (0 \leqslant r < 1). \tag{24}$$

On replacing r by r/R and φ by $(t - \theta)$, we finally get the following expression for $U(r, \theta)$:

$$U(r, \theta) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) \frac{R^2 - r^2}{R^2 - 2rR\cos(t-\theta) + r^2} \, dt, \tag{25}$$

It may be pointed out that if we wrote $(+k^2)$ instead of $(-k^2)$ for both sides of equation (18_1), we should have $(Ae^{k\theta} + Be^{-k\theta})$ instead of $(A \cos k\theta + B \sin k\theta)$ in expression (19), and the exponential expression is not periodic for any real k.

We have assumed in deducing formula (25) that a solution of the Dirichlet problem, i.e. a function $U(r, \theta)$, exists. Moreover, we have made use of the Fourier expansion (22) of $f(\theta)$, which is not necessarily valid, and we have substituted $r = R$ directly in this expansion. All this obliges us to check formula (25), i.e. we have to show that the integral on the right-hand side yields a harmonic function inside the circle $r < R$ and that $f(\theta)$ gives the boundary values of this function on the circumference. We generally refer to the integral in (25) as a Poisson integral.

196. Poisson integrals. We shall take the radius R of the circle as unity in this section so as to simplify the writing; in this case, (25) takes the form

$$U(r, \theta) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t) \frac{1 - r^2}{1 - 2r\cos(t - \theta) + r^2} \, dt. \tag{26}$$

The integral yields a function of r and θ, since the second factor of its integrand,

$$\frac{1 - r^2}{1 - 2r\cos(t - \theta) + r^2} \tag{27}$$

contains the parameters r and θ in addition to the variable of integration t. This second factor has a period of 2π with respect to θ, and the same is therefore true of function (26). It follows from the obvious inequality $1 - 2r\cos(t - \theta) + r^2 \geqslant 1 - 2r + r^2 = (1 - r)^2$ that (27) and its derivatives of any order are continuous functions of r and θ for $0 \leqslant r < 1$. This means that the integral of (26) can be differentiated with respect to r and θ under the integral sign [80], and this can only concern the factor (27). But it is easily shown by using the expression for Laplace's operator in polar coordinates [119] that (27) satisfies Laplace's equation. It follows at once from this that expression (26) defines a function $U(r, \theta)$ which is harmonic for $r < 1$. An important part of the proof still remains: to show that $U(r, \theta)$ is equal to $f(\theta)$ on the circumference $r = 1$.

We remark first of all that the harmonic function $U(r, \theta)$ is identically equal to unity if we set $f(t) = 1$ in (26), in other words:

$$1 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \frac{1 - r^2}{1 - 2r\cos(t - \theta) + r^2} \, dt. \tag{28}$$

The proof is as follows. We have from (24):

$$\frac{1-r^2}{1-2r\cos(t-\theta)+r^2} = 1 + 2\sum_{n=1}^{\infty} r^n \cos n(t-\theta) \quad (0 \leqslant r < 1),$$

the series on the right being uniformly convergent in regard to t, since its terms do not exceed $2r^n$ in absolute value. On integrating term by term with respect to t, we obtain (28).

Function $f(t)$ is defined on the circumference $r = 1$ and has period 2π, i.e. $f(-\pi) = f(\pi)$. We continue it outside the interval $(-\pi, \pi)$ in a periodic manner so as to get a continuous function $f(t)$ of period 2π in the interval $-\infty < t < +\infty$.

We introduce a new variable of integration $\varphi = t - \theta$ in place of t, so that $t = \varphi + \theta$ and $\mathrm{d}t = \mathrm{d}\varphi$. On taking account of the periodicity of $f(t)$ and $\cos(t-\theta)$, we can retain the original interval of integration $(-\pi, \pi)$ [142] and write:

$$U(r, \theta) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(\varphi+\theta)\frac{1-r^2}{1-2r\cos\varphi+r^2}\,\mathrm{d}\varphi. \tag{29}$$

Let the point (r, θ) tend to $(1, \theta_0)$ on the circumference. We have to show that, with this,

$$\lim U(r, \theta) = f(\theta_0).$$

We carry out the change of variable in the integral of (28), multiply both sides of the equation by $f(\theta_0)$, and subtract the equation obtained from equation (29):

$$U(r, \theta) - f(\theta_0) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} [f(\varphi+\theta) - f(\theta_0)]\frac{1-r^2}{1-2r\cos\varphi+r^2}\,\mathrm{d}\varphi. \tag{30}$$

We now want to show that the integral on the right tends to zero as $r \to 1$ and $\theta \to \theta_0$, i.e. that its absolute value is as small as desired for r sufficiently close to unity and θ to θ_0. For any given positive ε, an η can be found such that in the interval $-\eta < \varphi < \eta$,

$$|f(\varphi+\theta) - f(\theta_0)| < \frac{\varepsilon}{2} \tag{31}$$

for all θ sufficiently close to θ_0. We divide the interval of integration in (30) into three parts:

$$(-\pi, -\eta), (-\eta, \eta), (\eta, \pi). \tag{32}$$

We consider the absolute value of the integral over the second part:

$$I_2 = \frac{1}{2\pi} \int_{-\eta}^{+\eta} [f(\varphi + \theta) - f(\theta_0)] \frac{1 - r^2}{1 - 2r\cos\varphi + r^2} d\varphi.$$

On noticing that the fraction in the integrand is positive, replacing the difference term by its absolute value, and using (31), we get:

$$|I_2| < \frac{\varepsilon}{2} \cdot \frac{1}{2\pi} \int_{-\eta}^{+\eta} \frac{1 - r^2}{1 - 2r\cos\varphi + r^2} d\varphi,$$

or, on widening the interval of integration:

$$|I_2| < \frac{\varepsilon}{2} \cdot \frac{1}{2\pi} \int_{-\pi}^{+\pi} \frac{1 - r^2}{1 - 2r\cos\varphi + r^2} d\varphi,$$

whence, in view of (28):

$$|I_2| < \frac{\varepsilon}{2}. \tag{33}$$

We now consider the integral over the first of intervals (32). We have $\cos\varphi \leqslant \cos\eta$ in this interval, so that

$$1 - 2r\cos\varphi + r^2 \geqslant 1 - 2r\cos\eta + r^2 = (1 - r)^2 + 2r(1 - \cos\eta)$$

or

$$1 - 2r\cos\varphi + r^2 \geqslant 4r\sin^2\frac{\eta}{2}.$$

The absolute value of the difference $[f(\varphi + \theta) - f(\theta_0)]$ does not exceed some fixed positive number M, since $f(t)$ is a continuous periodic function. We can thus write for the integral over the first interval:

$$|I_1| < \frac{M}{8\pi r\sin^2\frac{\eta}{2}} (1 - r^2)(\pi - \eta),$$

whilst a similar inequality is obtained for the integral over the third of intervals (32). The right-hand side of this inequality tends to zero as $r \to 1$, so that the sum of the integrals over the first and third intervals has an absolute value less than $\varepsilon/2$ for all r sufficiently close to unity. If we take into account (33) and the arbitrary smallness of ε, we can say that the right-hand side of equation (30) in fact tends to zero as $r \to 1$ and $\theta \to \theta_0$.

The connection between the integral expression (26) and the Fourier series for $f(\theta)$ should be noticed. The Fourier series is of the form (22), where for the moment we put $R = 1$:

$$\frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos n\theta + B_n \sin n\theta), \tag{34}$$

the coefficients being defined by (23) with $R = 1$. If, for instance, $f(\theta)$ satisfies Dirichlet conditions [143], series (34) is convergent for all θ. We cannot assert this in the general case of a continuous function. We can always say, however, that A_n and $B_n \to 0$ as $n \to \infty$, so that the series

$$\frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos n\theta + B_n \sin n\theta)\, r^n \tag{35}$$

converges for $r < 1$, and as is clear from [195], the sum of the series in fact gives function (26). It also turns out that the sum of series (35) tends to $f(\theta)$ as $r \to 1$, i.e. the sum tends to the function from which the Fourier series (34) originated, although this latter series may in fact be divergent.

We apply this idea to any series, say

$$\sum_{n=1}^{\infty} u_n. \tag{36}$$

If this series is convergent to the sum s, Abel's theorems regarding power series [I, 148] show that the series

$$\omega(r) = \sum_{n=0}^{\infty} u_n r^n \tag{37}$$

is convergent for $0 \leqslant r < 1$, and in view of its uniform convergence in the interval $0 \leqslant r \leqslant 1$ [I, 149], we have:

$$\lim_{r \to 1-0} \omega(r) = s. \tag{38}$$

It may happen, however, that series (36) is divergent, whilst series (37) converges for $0 \leqslant r < 1$ and $\omega(r)$ has a limit as $r \to 1 - 0$, i.e. (38) is valid. In this case, s is called the generalized sum of divergent series (36) in Abel's sense, and we refer to an Abelian summation of (36). It follows at once from what has been said that the generalized sum exists for convergent series and is the same as the ordinary sum.

The results obtained above in regard to Poisson's integral can be stated thus: *the Fourier series of a continuous periodic function* $f(\theta)$

has an Abelian summation for any θ and a generalized sum equal to $f(\theta)$. It should be noticed that, in investigating the Poisson integral, the point (r, θ) has tended to $(1, \theta_0)$ in any manner, and not necessarily along a radius.

Let us suppose that $r > 1$ in integral (26). We may verify, exactly as above, that (26) yields a harmonic function outside the circle $r = 1$. We investigate its boundary values by rewriting it in the form:

$$U(r, \theta) = -\frac{1}{\pi}\int_{-\pi}^{+\pi} f(t) \frac{1 - \left(\frac{1}{r}\right)^2}{1 - 2\frac{1}{r}\cos(t-\theta) + \left(\frac{1}{r}\right)^2}\, dt. \qquad (26_1)$$

This new integral is the same as (26), provided we replace the r in the latter by $1/r$, whilst we have $1/r < 1$ since $r > 1$. Thus all our previous discussion is applicable, on changing r to $1/r$, to the integral (26_1), which therefore tends to $f(\theta_0)$ as the point (r, θ) tends to $(1, \theta_0)$ from outside the circle. Hence we can say that the function

$$V(r, \theta) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(t) \frac{r^2 - 1}{1 - 2r\cos(t-\theta) + r^2}\, dt$$

gives the solution of Dirichlet's problem with boundary values $f(\theta)$ for the part of the plane lying outside the circle $r = 1$. As is evident from the last expression, the function $V(r, \theta)$ has a finite limit when the point (r, θ) becomes infinitely remote:

$$\lim_{r\to\infty} V(r, \theta) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(t)\, dt.$$

As we remarked above, the solution $U(M)$ of the Dirichlet problem for an infinite part of the plane lying outside a closed contour l is unique if we assume that the required function tends to a finite limit when the point M is at an infinite distance (cf. Volume IV).

197. Dirichlet's problem for a sphere. Let R be the radius of a sphere (Σ) and let $f(M')$ be the given boundary value of a harmonic function on the surface of the sphere, M' being a variable point of the surface. We assume that the function $f(M')$ is continuous on the surface of the sphere.

Let r denote the distance of a variable point M of space from an arbitrary fixed point M_0 inside (Σ). In addition to M_0, we take the

point M_1 lying on the produced radius OM_0 of the sphere and such that (Fig. 142):

$$\overline{OM}_0 \cdot \overline{OM}_1 = R^2. \tag{39}$$

The point M_1 outside the sphere is sometimes said to be symmetrical to M_0 with respect to (Σ). Let r_1 denote the distance of the variable point M from M_1. If M lies at M' on the surface of (Σ), r and r_1 are connected by a simple relationship which we shall now deduce. The triangles OM_0M' and OM_1M' are similar, since they have a common angle with vertex at O and the sides forming this angle are proportional by (39). It follows from their similarity that

$$\frac{|M_0M'|}{|M_1M'|} = \frac{|OM_0|}{|OM'|} \quad \text{or} \quad \frac{r}{r_1} = \frac{|OM_0|}{R}$$

whence

$$\frac{1}{r_1} = \frac{\varrho}{R} \cdot \frac{1}{r}, \tag{40}$$

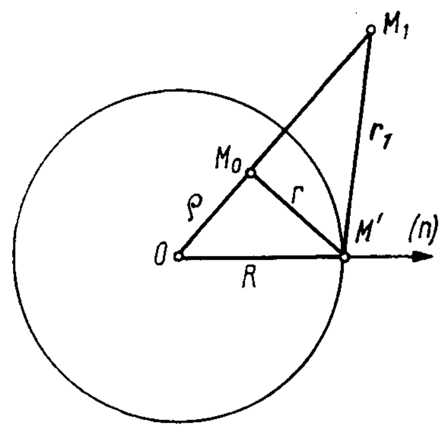

FIG. 142

where $\varrho = |OM_0|$ is the length of the radius vector from the centre of the sphere to M_0. Since M_1 lies outside the sphere, $1/r_1$ does not become infinite inside the sphere and therefore represents a harmonic function inside it [119]. Equation (40) gives the boundary values of this function on the surface of the sphere. Let $U(M)$ be the solution of the Dirichlet problem. Equation (13) gives:

$$U(M_0) = \frac{1}{4\pi} \iint\limits_{(\Sigma)} \left(\frac{1}{r} \frac{\partial U}{\partial n} - U \frac{\partial \frac{1}{r}}{\partial n} \right) \mathrm{d}S. \tag{41}$$

On the other hand, we find on applying expression (6) to the harmonic functions U and $V = 1/r_1$:

$$0 = \iint\limits_{(\Sigma)} \left(\frac{1}{r_1} \frac{\partial U}{\partial n} - U \frac{\partial \frac{1}{r_1}}{\partial n} \right) \mathrm{d}S. \tag{42}$$

On multiplying both sides of (42) by the constant $R/4\pi\varrho$ and subtracting from (41), we can eliminate $\partial U/\partial n$ by using (40):

$$U(M_0) = \frac{1}{4\pi} \iint\limits_{(\Sigma)} U \cdot \left[\frac{R}{\varrho} \frac{\partial \left(\frac{1}{r_1}\right)}{\partial n} - \frac{\partial \left(\frac{1}{r}\right)}{\partial n} \right] \mathrm{d}S.$$

The given function $f(M')$ represents the values of U on (Σ), so that we can write the above as

$$U(M_0) = \frac{1}{4\pi} \iint\limits_{(\Sigma)} f(M') \left[\frac{R}{\varrho} \frac{\partial \left(\frac{1}{r_1} \right)}{\partial n} - \frac{\partial \left(\frac{1}{r} \right)}{\partial n} \right] dS. \tag{43}$$

This is in fact the solution of the Dirichlet problem for a sphere since the integrand is known. We shall put the difference in square brackets in another form. We first remark that the surfaces $r = \text{const.}$ are spheres with centre M_0, so that grad r is a unit vector in the direction $\overline{M_0 M}$, and consequently,

$$\frac{\partial r}{\partial n} = \text{grad}_n r = \cos(r, n)$$

and

$$\frac{\partial \left(\frac{1}{r} \right)}{\partial n} = \frac{\partial \left(\frac{1}{r} \right)}{\partial r} \cdot \frac{\partial r}{\partial n} = -\frac{1}{r^2} \cos(r, n).$$

Similarly,

$$\frac{\partial \left(\frac{1}{r_1} \right)}{\partial n} = -\frac{1}{r_1^2} \cos(r_1, n),$$

where the r and r_1 in the cosines denote the directions $\overline{M_0 M}$ and $\overline{M_1 M}$. This gives:

$$\frac{R}{\varrho} \frac{\partial \left(\frac{1}{r_1} \right)}{\partial n} - \frac{\partial \left(\frac{1}{r} \right)}{\partial n} = \frac{1}{r^2} \cos(r, n) - \frac{R}{\varrho r_1^2} \cos(r_1, n). \tag{44}$$

On introducing $\varrho_1 = |OM_1| = R^2/\varrho$, we can write from triangles $OM'M_0$ and $OM'M_1$:

$$\varrho^2 = R^2 + r^2 - 2Rr\cos(r, n); \quad \varrho_1^2 = R^2 + r_1^2 - 2Rr_1 \cos(r_1, n).$$

Having thus determined $\cos(r, n)$ and $\cos(r_1, n)$, substitution in equation (44) gives us, in view of (40) and the definition of ϱ_1:

$$\frac{R}{\varrho} \frac{\partial \left(\frac{1}{r_1} \right)}{\partial n} - \frac{\partial \left(\frac{1}{r} \right)}{\partial n} = \frac{R^2 - \varrho^2}{Rr^3},$$

so that we can now write (43) as:

$$U(M_0) = \frac{1}{4\pi R} \iint\limits_{(\Sigma)} f(M') \frac{R^2 - \varrho^2}{r^3} dS, \tag{45}$$

Alternatively, if we bring in the angle γ formed by the radius vector $\overline{OM}_0$ with the variable radius vector $\overline{OM'}$, the angular spherical coordinates (θ', φ') of the point M', and the spherical coordinates $(\varrho_0, \theta_0, \varphi_0)$ of M_0 with the origin at O, we can write:

$$U(\varrho, \theta_0, \varphi_0) = \frac{R}{4\pi} \int_0^{2\pi} \int_0^{\pi} f(\theta', \varphi') \frac{R^2 - \varrho^2}{(R^2 - 2\varrho R \cos\gamma + \varrho^2)^{3/2}} \sin\theta' \, d\theta' \, d\varphi'. \quad (46)$$

This integral form of $U(M_0)$ is analogous to the Poisson integral obtained in the case of a plane. To prove that the integral appearing in (45) yields a harmonic function, it is sufficient to show that the fraction $(R^2 - \varrho^2)/r^3$ is a harmonic function of M_0 for fixed M'. We introduce spherical coordinates with the origin at M' and the z axis directed from M' to O and with the usual notation of $\theta = \angle OM'M_0$. We now have:

$$\frac{R^2 - \varrho^2}{r^3} = \frac{2R\cos\theta}{r^2} - \frac{1}{r}.$$

We find on substituting the right-hand side above in the spherical coordinate form of Laplace's equation that we in fact have a harmonic function of the point M_0. We now show that, for any position of M_0 inside the sphere:

$$\frac{1}{4\pi R} \iint_{\Sigma} \frac{R^2 - \varrho^2}{r^3} \, dS = 1. \quad (*)$$

We take spherical coordinates with the origin at O and the z axis directed from O to M_0, with in this case $\theta = \angle M_0 OM'$ and $r^2 = R^2 - 2R\,\varrho\cos\theta + \varrho^2$. The integral in equation (*) becomes

$$\frac{R^2 - \varrho^2}{4\pi R} \int_0^{2\pi} \int_0^{\pi} \frac{R^2 \sin\theta \, d\theta \, d\varphi}{(R^2 - 2R\varrho\cos\theta + \varrho^2)^{3/2}} =$$

$$= \frac{(R^2 - \varrho^2) R}{2} \int_0^{\pi} \frac{\sin\theta \, d\theta}{(R^2 - 2\varrho R\cos\theta + \varrho^2)^{3/2}} =$$

$$= \frac{R^2 - \varrho^2}{2\varrho} (R^2 - 2\varrho R\cos\theta + \varrho^2)^{-1/2} \Big|_{\theta=\pi}^{\theta=0}$$

whence, if we take into account the fact that $\varrho < R$, we get expression (*):

$$\frac{1}{4\pi R} \iint_{\Sigma} \frac{R^2 - \varrho^2}{r^3} \, dS = \frac{R^2 - \varrho^2}{2\varrho} \left(\frac{1}{R - \varrho} - \frac{R}{R + \varrho} \right) = 1.$$

The further proof of the fact that integral (45) has boundary values $f(M)$ on the sphere proceeds as in the case of the Poisson integral.

The solution of the exterior Dirichlet problem with boundary values $f(M')$ is given by:

$$U(M_0) = \frac{1}{4\pi R} \iint_{\Sigma'} f(M') \frac{\varrho^2 - R^2}{r^3} \, dS \tag{45_1}$$

or

$$U(\varrho, \theta_0, \varphi_0) = \frac{R}{4\pi} \int_0^{2\pi} \int_0^{\pi} f(\theta', \varphi') \frac{\varrho^2 - R^2}{(R^2 - 2\varrho R \cos\gamma + \varrho^2)^{3/2}} \sin\theta' \, d\theta' \, d\varphi', \tag{46_1}$$

where $\varrho = |\overline{OM}_0|$, $r = |\overline{M_0 M'}|$ and $\gamma = \angle M_0 OM'$, whilst here $\varrho > R$. It may be shown as above that the integral of (45_1) yields a harmonic function outside the sphere. To verify that $U(M_0)$ has the boundary values $f(M)$, we rewrite (46_1) as:

$$U(\varrho, \theta_0, \varphi_0) =$$
$$= \frac{\varrho'}{4\pi} \int_0^{2\pi} \int_0^{\pi} f(\theta', \varphi') \frac{R'^2 - \varrho'^2}{(R'^2 - 2\varrho' R' \cos\gamma + \varrho'^2)^{3/2}} \sin\theta' \, d\theta' \, d\varphi', \tag{46_2}$$

where $\varrho' = \varrho^{-1}$ and $R' = R^{-1}$. With this, we have $\varrho' < R'$, and when the point $(\varrho, \theta_0, \varphi_0)$ tends to $M(R, \theta, \varphi)$, lying on the sphere Σ, $(\varrho', \theta_0, \varphi_0)$ tends to (R', θ, φ). We have, from the result obtained for the interior of the sphere:

$$\frac{R'}{4\pi} \int_0^{2\pi} \int_0^{\pi} f(\theta', \varphi') \frac{R'^2 - \varrho'^2}{(R'^2 - 2\varrho' R' \cos\gamma + \varrho'^2)^{3/2}} \sin\theta' \, d\theta' \, d\varphi' \to f(M),$$

and we can say, since $\varrho' \to R'$, that the right-hand side of (46_2) in fact tends to $f(M)$, as we wished to prove. It should also be noticed that, by (46_1), $U(\varrho, \theta_0, \varphi_0)$ tends to zero when M_0 is at an infinite distance, i.e. when $\varrho \to \infty$. This follows from the fact that the numerator in the integrand of (46_1) contains ϱ^2, whilst the denominator is clearly of order ϱ^3.

198. Green's function. The solution of the general case of the interior Dirichlet problem for any surface (S) may be deduced from the above solution for the spherical case. Expression (13) does not give the solution directly since $\partial U/\partial n$ appears under the sign of the double integral, in addition to U itself, the values of which are known on the surface. To solve the problem, we have to eliminate the derivative. Let M_0 be a fixed point inside (S). Let $G_1(M; M_0)$ be a known function with the following two properties: (1) considered as a

function of the variable point M, it is harmonic inside (S); (2) its boundary value on (S) is $1/r$, where r is the distance from the variable point of (S) to M_0. Let $U(M)$ be the required solution of the Dirichlet problem. We can write, on applying (6) to the harmonic functions $U(M)$ and $G_1(M; M_0)$:

$$0 = \iint\limits_{(S)} \left[U(M) \frac{\partial G_1(M;\ M_0)}{\partial n} - G_1(M;\ M_0) \frac{\partial U(M)}{\partial n} \right] \mathrm{d}S,$$

or in view of the boundary value of $G_1(M; M_0)$:

$$0 = \iint\limits_{(S)} \left[U(M) \frac{\partial G_1(M;\ M_0)}{\partial n} - \frac{1}{r} \frac{\partial U(M)}{\partial n} \right] \mathrm{d}S.$$

On multiplying this equation by $\pi/4$ and adding to (13), we get:

$$U(M_0) = -\frac{1}{4\pi} \iint\limits_{(S)} U(M) \frac{\partial}{\partial n} \left[\frac{1}{r} - G_1(M;\ M_0) \right] \mathrm{d}S. \tag{47}$$

This gives the solution of the Dirichlet problem, if $G_1(M; M_0)$ is known. The difference in square brackets:

$$G(M;\ M_0) = \frac{1}{r} - G_1(M;\ M_0), \tag{48}$$

is known as *the Green function for the domain bounded by (S) with pole at M_0*. Two basic properties of Green functions follow from the definition of $G_1(M; M_0)$:

1. $G(M; M_0)$ is harmonic inside (S) except at the point M_0 where it becomes infinite; whereas the difference $G(M; M_0) - 1/r$ remains finite and is harmonic everywhere inside (S).

2. $G(M; M_0)$ has zero boundary values on (S).

If we place a unit positive electric charge at M_0 and look on (S) as a conducting surface joined to earth, Green's function $G(M; M_0)$ gives the electrostatic potential of the field obtained inside (S).

In the case of a sphere, $G_1(M; M_0)$ is equal to $(R/\varrho) \cdot 1/r_1$ from (40), and Green's function becomes:

$$G(M;\ M_0) = \frac{1}{r} - \frac{R}{\varrho} \cdot \frac{1}{r_1}. \tag{49}$$

We have obtained (47) by using (13) and applying Green's integral formula to $U(M)$ and $G_1(M; M_0)$. This procedure requires special proof, involving an investigation of the behaviour of the derivatives on approaching the surface (S). A rigorous proof of (47) with wide assumptions regarding the surface (S) and the function $U(M)$ on (S) was first given by A. M. Lyapunov.

We have a precisely analogous formula for the solution of the interior Dirichlet problem in the plane case:

$$U(M_0) = -\frac{1}{2\pi} \int\limits_{l} U(M) \frac{\partial G(M;\ M_0)}{\partial n} \mathrm{d}s, \tag{48$_1$}$$

where the Green function $G(M; M_0)$ for the domain with contour l and pole at M_0 must have the following two properties:

1. $G(M; M_0)$ is harmonic inside (l) except at M_0, where it becomes infinite; whereas the difference $G(M; M_0) - \log 1/r$ is harmonic at all interior points including M_0.

2. $G(M; M_0)$ is zero on the contour (l).

It may readily be seen that only one function can exist with these two properties. If there were two, say $G^{(2)}(M; M_0)$ and $G^{(1)}(M; M_0)$, their difference would be harmonic everywhere inside S or l and would be zero on these boundaries, i.e. it would be identically zero inside S or l.

199. The case of a half-space. As an example of the application of (47), we consider the Dirichlet problem for a half-space. We require to find the function $U(x, y, z)$ which is harmonic in the half-space $z > 0$ and has known boundary values $f(x, y)$ on the plane $z = 0$:

$$U|_{z=0} = f(x, y). \tag{50}$$

Let r be the distance of a variable point M from the point $M_0(x_0, y_0, z_0)$, where $z_0 > 0$, and r_1 be the distance of M from $M_0(x_0, y_0, -z_0)$, the symmetrical point to M_0 with respect to the plane $z = 0$. The fraction $1/r_1$ is a harmonic function of M in the half-space $z > 0$, since M_0 lies outside this region. If M lies on the plane $z = 0$, obviously, $1/r_1 = 1/r$. Green's function becomes here:

$$G(M;\ M_0) = \frac{1}{r} - \frac{1}{r_1} =$$

$$= \frac{1}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2}} - \frac{1}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z+z_0)^2}}.$$

From the point of view of the half-space, the outward normal to the plane $z = 0$ is in the opposite direction to the z axis, i.e. $\partial/\partial n = -\partial/\partial z$, and (47) gives:

$$U(x_0, y_0, z_0) = \frac{1}{4\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \frac{\partial}{\partial z} \left[\frac{1}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2}} - \right.$$

$$\left. - \frac{1}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z+z_0)^2}} \right]_{z=0} dx\, dy.$$

We have to put $z = 0$ after differentiating the square bracket. Some simple working eventually gives us:

$$U(x_0, x_0, z_0) = \frac{z_0}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{f(x, y)}{[(x-x_0)^2 + (y-y_0)^2 + z_0^2]^{3/2}}\, dx\, dy. \tag{51}$$

We shall not verify that the right-hand side represents a harmonic function and has limiting values $f(x, y)$ when (x_0, y_0, z_0) tends to $(x, y, 0)$. Infinitely remote points lie on the surface of the domain in the present case, and it may readily be seen that the solution has the following property: if $f(x, y)$ is continuous at infinity, i.e. has a definite finite limit a as the point (x, y) moves to infinity on the plane $z = 0$, $U(x_0, y_0, z_0)$ has the same limit a as (x_0, y_0, z_0) moves to an infinite distance in any manner in the half-plane $z > 0$.

In other words, our solution has the required boundary values even for infinitely remote points of the plane provided $f(x, y)$ is continuous at these points.

In the same way, solution of the Dirichlet problem for the half-plane $y > 0$ leads to a Green function of the form:

$$\log \frac{1}{r} - \log \frac{1}{r_1} = \log \frac{1}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} - \log \frac{1}{\sqrt{(x - x_0)^2 + (y + y_0)^2}} ,$$

and with boundary values

$$U|_{y=0} = f(x), \tag{52}$$

(48_1) gives the solution as

$$U(x_0, y_0) = \frac{y_0}{\pi} \int_{-\infty}^{+\infty} \frac{f(x)}{(x - x_0)^2 + y_0^2} \, dx. \tag{53}$$

The detailed treatment of Neumann's problem is left over to Volume IV.

200. Potential of a distributed mass. We take the non-homogeneous Laplace equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = \varphi(x, y, z) \tag{54}$$

in a finite domain D with a surface S. The general solution consists of the sum of some particular solution and a function harmonic in D. Let a solution exist to which (9) is applicable. Since the derivative of $1/r$ with respect to any fixed direction satisfies Laplace's equation, the integrand in the surface integral of (9) and the integral itself are harmonic in D. Thus the triple integral in (9) must satisfy equation (54). But by (54), the ΔU in the triple integral can be replaced by $\varphi(x, y, z)$, so that we have as a particular solution of (54):

$$U(x, y, z) = -\frac{1}{4\pi} \iiint_D \frac{\varphi(\xi, \eta, \zeta)}{r} \, dv \tag{55}$$

$$r = \sqrt{(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2} .$$

We have obtained this result on the assumption that (54) has a solution to which (9) is applicable. A full solution of the problem requires a more detailed consideration of the volume potential (55) with definite assumptions regarding $\varphi(N)$. We write $\mu(N) = -\varphi(N)/4\pi$ and consider the following distributed mass potential:

$$V(M) = \iiint_D \frac{\mu(N)}{r} \, dv \tag{56}$$

or

$$V(x, y, z) = \iiint_D \frac{\mu(\xi, \eta, \zeta)}{r} \, dv. \tag{56$_1$}$$

Let $\mu(N)$ be continuous in D as far as S. As already remarked, integral (56) is a particular solution if M lies outside D. In this case, $V(M)$ has partial derivatives of any order. These derivatives can be obtained by differentiating under the integral sign, and $V(M)$ satisfies Laplace's equation $\Delta V = 0$. If M belongs to D, (56) exists as an improper integral and the integral obtained by differentiating the integrand e.g. with respect to x likewise exists. We have not shown, however, that this latter integral yields the partial derivative of V with respect to x. We shall prove two theorems regarding integral (56):

THEOREM 1. *If $\mu(N)$ is continuous in D as far as S, $V(M)$ and its first order partial derivatives are continuous throughout space and the derivatives can be obtained by differentiation under the integral sign.*

We shall carry out the proof for any position of M in regard to D. We replace $1/r$ by a new function which only differs from $1/r$ for $r < \varepsilon$, where ε is a given positive number, and which is continuous and has continuous derivatives with respect to the coordinates as far as $r = 0$. We do this by taking, for $r < \varepsilon$, the polynomial: $a + \beta r^2 = a + \beta [(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2]$, a and β being chosen so that, for $r = \varepsilon$:

$$a + \beta\varepsilon^2 = \frac{1}{\varepsilon} \text{ and } 2\beta\varepsilon = -\frac{1}{\varepsilon^2};$$

in this way, we obtain continuity of the derivatives at the meeting-point of functions $1/r$ and $a + \beta r^2$, i.e. at $r = \varepsilon$. The equations written give: $a = 3/2\varepsilon$, $\beta = -1/2\varepsilon^3$, and we arrive at the function $g_\varepsilon(r)$, defined by the equations:

$$\begin{aligned} g_\varepsilon(r) &= \frac{1}{r} \quad \text{for} \quad r \geqslant \varepsilon \\ g_\varepsilon(r) &= \frac{3}{2\varepsilon} - \frac{1}{2\varepsilon^3} r^2 \quad \text{for} \quad r < \varepsilon. \end{aligned} \tag{57}$$

On replacing the $1/r$ in integral (56) by this function, we get instead of $V(M)$:

$$V_\varepsilon(M) = \iiint_D \mu(N) \, g_\varepsilon(r) \, dv, \tag{58}$$

which is continuous throughout space and has continuous partial derivatives which can be found by differentiating under the integral

sign, since the integrand of (58) is itself continuous and has continuous derivatives for $r \geqslant 0$. We can write, for instance:

$$\frac{\partial V_\varepsilon(M)}{\partial x} = \iiint_D \mu(N) \frac{\partial}{\partial x} g_\varepsilon(r)\, \mathrm{d}v. \tag{59}$$

We consider the difference:

$$V(M) - V_\varepsilon(M) = \iiint_D \mu(N) \left[\frac{1}{r} - g_\varepsilon(r)\right] \mathrm{d}v \tag{60}$$

Since $1/r$ and $g_\varepsilon(r)$ are identical for $r \geqslant \varepsilon$, their difference appearing on the right-hand side is zero for all points N lying outside the sphere σ_ε with centre M and radius ε. If, for instance, M lies outside D and ε is less than the distance of M from D, the integral on the right-hand side of (60) is zero.

In other cases, the sphere σ_ε can belong partly or wholly in D. If we write m for the greatest absolute value of $\mu(N)$ in D and remember that $g_\varepsilon(r)$ is a positive function, we can write for the integrand in (60):

$$\left| \mu(N) \left[\frac{1}{r} - g_\varepsilon(r)\right] \right| < m \left[\frac{1}{r} + g_\varepsilon(r)\right], \tag{61}$$

whilst the integrand vanishes outside σ_ε, as already remarked. We can clearly write, on integrating the positive function on the right-hand side of (61) over the whole of σ_ε:

$$| V(M) - V_\varepsilon(M) | \leqslant m \int_0^\varepsilon \int_0^{2\pi} \int_0^\pi \left[\frac{1}{r} + g_\varepsilon(r)\right] r^2 \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi \, \mathrm{d}r\,.$$

We obtain on replacing $g_\varepsilon(r)$ by the second of expressions (57) and carrying out the quadratures:

$$| V(M) - V_\varepsilon(M) | < \frac{18\pi}{5} m \varepsilon^2.$$

It is clear from this that the continuous functions $V_\varepsilon(M)$ tend uniformly as regards the position of the point M to $V(M)$ as $\varepsilon \to 0$, so that $V(M)$ is likewise a continuous function [I, 144]. We investigate the partial derivatives of $V(M)$ by differentiating the integral of (56) with respect to x under the integral sign and writing $W(M)$ for the new integral obtained:

$$W(M) = \iiint_D \mu(N) \frac{\partial}{\partial x} \left(\frac{1}{r}\right) \mathrm{d}v. \tag{62}$$

As above, we now consider the difference:

$$W(M) - \frac{\partial V_\varepsilon(M)}{\partial x} = \iiint\limits_D \mu(N)\left[\frac{\partial}{\partial x}\left(\frac{1}{r}\right) - \frac{\partial}{\partial x} g_\varepsilon(r)\right] \mathrm{d}v$$

On noticing that we have, for any function $h(r)$:

$$\frac{\partial}{\partial x} h(r) = \frac{\mathrm{d}h(r)}{\mathrm{d}r} \frac{x-\xi}{r}$$

and that $|(x-\xi/r)| \leqslant 1$, we can write as regards the last integrand:

$$\left|\mu(N)\left[\frac{\partial}{\partial x}\left(\frac{1}{r}\right) - \frac{\partial}{\partial x} g_\varepsilon(r)\right]\right| \leqslant m\left[\frac{1}{r^2} + \left|\frac{\mathrm{d}g_\varepsilon(r)}{\mathrm{d}r}\right|\right],$$

or, precisely as above:

$$\left|W(M) - \frac{\partial V_\varepsilon(M)}{\partial x}\right| \leqslant m \int\limits_0^\varepsilon \int\limits_0^{2\pi} \int\limits_0^\pi \left[\frac{1}{r^2} + \left|\frac{\mathrm{d}g_\varepsilon(r)}{\mathrm{d}r}\right|\right] r^2 \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\varphi \,\mathrm{d}r.$$

Since, by (57),

$$\left|\frac{\mathrm{d}g_\varepsilon(r)}{\mathrm{d}r}\right| = \frac{r}{\varepsilon^3} \quad \text{for } r \leqslant \varepsilon,$$

we obtain after carrying out the quadratures:

$$\left|W(M) - \frac{\partial V_\varepsilon(M)}{\partial x}\right| \leqslant 5\pi m\varepsilon,$$

whence it follows that as $\varepsilon \to 0$, $\partial V_\varepsilon(M)/\partial x \to W(M)$ uniformly with respect to M. We proved above that $V_\varepsilon(M)$ tends uniformly to $V(M)$. Remembering the theorem of [I, 144], we see that $W(M)$ is the partial derivative of $V(M)$ with respect to x, i.e. using (62):

$$\frac{\partial}{\partial x} \iiint\limits_D \mu(N) \frac{1}{r} \mathrm{d}v = \iiint\limits_D \mu(N) \frac{\partial}{\partial x}\left(\frac{1}{r}\right) \mathrm{d}v.$$

The continuity of $W(M)$ follows from the continuity of partial derivatives (59) and their uniform convergence to $W(M)$, so that the theorem is fully proved. The derivatives with respect to y and z may be investigated in precisely the same way. It may be noticed that our proof has made use only of the boundedness and integrability of $\mu(N)$.

201. Poisson's equation. We must amplify our assumptions regarding $\mu(N)$ before finding the second order derivatives of $V(M)$.

THEOREM 2. *If the continuous function $\mu(N)$ has continuous first order derivatives inside D, $V(M)$ has continuous second order derivatives*

inside D and satisfies the equation inside D:

$$\Delta V(M) = -4\pi\mu(M). \tag{63}$$

We take any fixed point $M_0(x_0, y_0, z_0)$ inside D. Let σ_ε be a sphere with centre M_0 and radius ε lying inside D, and let D_1 be the part of D lying outside σ_ε. We split potential (56) into two terms:

$$V(M) = \iiint\limits_{D_1} \mu(N)\frac{1}{r}\,\mathrm{d}v + \\ + \iiint\limits_{\sigma_\varepsilon} \mu(N)\frac{1}{r}\,\mathrm{d}v = V_1(M) + V_0(M), \tag{64}$$

and, in view of Theorem 1:

$$\frac{\partial V(M)}{\partial x} = \iiint\limits_{D_1} \mu(N)\frac{\partial}{\partial x}\left(\frac{1}{r}\right)\mathrm{d}v + \\ + \iiint\limits_{\sigma_\varepsilon} \mu(N)\frac{\partial}{\partial x}\left(\frac{1}{r}\right)\mathrm{d}v = \frac{\partial V_1(M)}{\partial x} + \frac{\partial V_0(M)}{\partial x}. \tag{65}$$

We have:

$$\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = -\frac{\partial}{\partial \xi}\left(\frac{1}{r}\right);\quad \left(r = \sqrt{(\xi - x)^2 + (\eta - y)^2 + (\zeta - z)^2}\right)$$

and we can therefore write:

$$\mu(N)\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = -\frac{\partial}{\partial \xi}\left[\mu(M)\frac{1}{r}\right] + \frac{\partial \mu(N)}{\partial \xi}\cdot\frac{1}{r}.$$

On substituting this expression for the integrand of the integral over σ_ε in (65) and applying Ostrogradskii's formula we get:

$$\frac{\partial V(M)}{\partial x} = \iiint\limits_{D_1} \mu(N)\frac{\partial}{\partial x}\left(\frac{1}{r}\right)\mathrm{d}v + \\ + \iiint\limits_{\sigma_\varepsilon} \frac{\partial \mu(N)}{\partial \xi}\cdot\frac{1}{r}\,\mathrm{d}v - \iint\limits_{S_\varepsilon} \mu(N)\cos(n, x)\frac{1}{r}\,\mathrm{d}S, \tag{66}$$

where S_ε is the surface of sphere σ_ε and n is the direction of the outward normal to S_ε at the point N. The first term on the right-hand side of (66) is a proper integral for M lying inside σ_ε and has derivatives of all orders inside σ_ε. The same can be said of the third term, consisting of an integral over the surface of σ_ε. The second term is the volume integral over σ_ε with continuous density $\partial\mu(N)/\partial\xi$, and by Theorem 1, has continuous first order derivatives throughout space. Hence we

can say that $\partial V(M)/\partial x$ has continuous first order derivatives inside σ_ε. In view of the arbitrariness of choice of M_0 in D, we can say that $\partial V(M)/\partial x$ has continuous first order derivatives everywhere inside D. On applying the same arguments to $\partial V(M)/\partial y$ and $\partial V(M)/\partial z$, it can be asserted that $V(M)$ has continuous second order derivatives inside D. It now remains for us to prove (63) for any point M_0 inside D.

We turn to equations (64) and (66). As we know, the potential $V_1(M)$ of the distributed mass over domain D_1 is a harmonic function inside σ_ε, since σ_ε lies outside D_1, i.e. $\Delta V_1(M) = 0$ inside σ_ε, so that $\Delta V(M) = \Delta V_0(M)$ inside σ_ε. It follows that $\Delta V(M)$ is simply obtained by first differentiating the terms of (66) with integration over σ_ε and S_ε with respect to x under the integral sign (using Theorem 1), then obtaining analogous expressions for the second order derivatives with respect to y and z, and finally adding all three derivatives. It must be borne in mind here that only the factor $1/r$ in the integrands depends on (x, y, z). Having thus found $\Delta V(M)$ inside σ_ε, we take its value at the centre M_0 of sphere σ_ε. If we write $\Delta V(M_0)$ for this value, and r_0 for the distance from M_0 to the variable point of integration, we get:

$$\Delta V(M_0) = \iiint\limits_{\sigma_\varepsilon} \left[\frac{\partial \mu(N)}{\partial \xi}\frac{\xi - x_0}{r_0^3} + \frac{\partial \mu(N)}{\partial \eta}\frac{\eta - y_0}{r_0^3} + \right.$$
$$\left. + \frac{\partial \mu(N)}{\partial \zeta}\frac{\zeta - z_0}{r_0^3}\right] dv - \iint\limits_{S_\varepsilon} \mu(N)\left[\frac{\xi - x_0}{r_0^3}\cos(n, x) + \right.$$
$$\left. + \frac{\eta - y_0}{r_0^3}\cos(n, y) + \frac{\zeta - z_0}{r_0^3}\cos(n, z)\right] dS. \tag{67}$$

This expression is valid for any choice of ε, provided the sphere σ_ε lies inside D, and the value of $\Delta V(M_0)$ is clearly independent of the choice of ε. We let ε tend to zero, and show that the triple integral likewise tends to zero. It is sufficient to take the integral of one of the terms. Let m be the greatest absolute value of the continuous function $\partial \mu(N)/\partial \xi$ in a sufficiently small fixed sphere σ_{ε_0}. Recalling that $|(\xi - x_0)/r_0| \leqslant 1$, we have for $\varepsilon \leqslant \varepsilon_0$:

$$\left|\iiint\limits_{\sigma_\varepsilon} \frac{\partial \mu(N)}{\partial \xi} \cdot \frac{\xi - x_0}{r_0^3}\, dv\right| \leqslant m \iiint\limits_{\sigma_\varepsilon} \frac{dv}{r_0^2}.$$

On introducing spherical coordinates with origin at M_0 and writing $dv = r_0^2 \sin\theta \, d\theta \, d\varphi \, dr_0$, it may be seen that the right-hand side is

equal to $m \cdot 4\pi\varepsilon$, whence it follows that the triple integral tends to zero as $\varepsilon \to 0$.

We now consider the surface integral of (67). Since the outward normal n is directed along the radius of the sphere, we have:

$$\frac{\xi - x_0}{r_0^3}\cos(n, x) + \frac{\eta - y_0}{r_0^3}\cos(n, y) + \frac{\zeta - z_0}{r_0^3}\cos(n, z) =$$
$$= \frac{1}{r_0^2}\left[\cos^2(n, x) + \cos^2(n, y) + \cos^2(n, z)\right] = \frac{1}{r_0^2},$$

so that the surface integral can be written as

$$\frac{1}{\varepsilon^2}\iint\limits_{S_\varepsilon} \mu(N)\, \mathrm{d}S,$$

or, on using the mean value theorem, as

$$\frac{1}{\varepsilon^2}\iint\limits_{S_\varepsilon} \mu(N)\, \mathrm{d}S = 4\pi\mu(N_\varepsilon),$$

where N_ε is a point on S_ε. As $\varepsilon \to 0$, N_ε tends to M_0 and $\mu(N_\varepsilon) \to$ $\to \mu(M_0)$, and the surface integral of (67) gives in the limit $4\pi\mu(M_0)$, which brings us to (63). We generally refer to (63) as Poisson's formula or Poisson's equation.

It follows at once from the theorem proved that, if $\varphi(x, y, z)$ is continuous in a domain D as far as its surface S and has continuous first order partial derivatives inside D, (55) gives the solution of equation (54). We remark that, if $\varphi(N)$ is defined throughout space and diminishes rapidly enough as N moves away to infinity, we can take D as the whole of space.

Similar theorems may be proved for the integral over a plane domain:

$$V(M) = \iint\limits_{B} \mu(N) \log\frac{1}{r}\, \mathrm{d}\sigma$$

or

$$V(x, y) = \iint\limits_{B} \mu(\xi, \eta) \log\frac{1}{r}\, \mathrm{d}\sigma; \quad \left(r = \sqrt{(\xi - x)^2 + (\eta - y)^2}\right).$$

If $\mu(N)$ is continuous in B as far as its contour, $V(M)$ is itself continuous and has continuous first order partial derivatives throughout the plane, these derivatives being obtainable by differentiation under the integral sign. If, moreover, $\mu(N)$ has continuous first order partial derivatives inside B, $V(M)$ has continuous second order partial

derivatives inside B and satisfies Poisson's equation at every interior point of B:

$$\Delta V(M) = -2\pi\mu(N)$$

In addition to integral (55), we consider

$$U_1(M) = -\frac{1}{4\pi}\iiint\limits_D \varphi(N)\, G(M\,;\,N)\,\mathrm{d}v\,, \tag{55$_1$}$$

where $G(M; N)$ is Green's function in domain D with pole N. The integration in (55) is carried out with respect to the point N. We can write, on taking into account expression (48):

$$U_1(M) = -\frac{1}{4\pi}\iiint\limits_D \frac{\varphi(N)}{r}\,\mathrm{d}v + \frac{1}{4\pi}\iiint\limits_D \varphi(N)\, G_1(M\,;\,N)\,\mathrm{d}v\,,$$

where $G_1(M; N)$ is a harmonic function of M everywhere inside D and has the boundary value $1/\varrho$ on S, ϱ being the distance from a variable point on S to the point N. Since M appears as a parameter in the integrand of the second integral whilst $G_1(M; N)$ is harmonic everywhere inside D, it follows that the second integral is likewise a harmonic function of M inside D. Laplace's operator yields $\varphi(M)$ from the first term on the right-hand side by what has been proved, so that the $U_1(M)$ defined by (55_1) satisfies equation (54). Furthermore, in view of the fact that $G(M; N)$ has zero boundary values on S, (55_1) shows us that $U_1(M)$ satisfies the boundary condition on S:

$$U_1(M)\,|_S = 0.$$

Equation (55_1) *defines the solution of equation* (54) *satisfying the boundary condition written.* The boundary values of solution (55), obtained as values of the integral on the right when the point (x, y, z) lies on S, depend on $\varphi(x, y, z)$. It should be remarked that the above investigation of function (55_1) is not strictly rigorous; we require a further investigation of the dependence of $G(M; N)$ on the point N, a proof that we can differentiate under the integral sign and can pass to the limit under the integral when M tends to a point on the surface S (cf. Vol. IV).

202. Kirchhoff's formula. Equation (13) gives the value at every interior point of a function harmonic inside a surface S as an integral over S. An analogous expression can be got for the function $V(x, y, z, t) = V(M; t)$, satisfying the wave equation

$$\frac{\partial^2 V}{\partial t^2} = a^2\,\Delta V\,. \tag{68}$$

Let $V(M; t)$ be continuous together with its derivatives to the second order in a domain D bounded by a surface S, with all $t > 0$. Let M_0 be a fixed point in D and let $r = M_0M$ denote the distance from M_0 to a variable point M. We apply general formula (9) to the function

$$U(x,y,z,t) = V\left(x,\, y,\, z,\, t - \frac{r}{a}\right), \tag{69}$$

or more briefly,

$$U(M, t) = V\left(M,\ t - \frac{r}{a}\right). \tag{70}$$

If $\omega(t)$ is a function of t, we write $[\omega]$ for the function obtained by replacing the t in $\omega(t)$ by $t - r/a$, i.e. $[\omega] = \omega\,(t - r/a)$.

We usually call $[\omega]$ *the retarded value of the function* $\omega(t)$. The meaning of the term becomes clear if we take a as the velocity of propagation of some process.

We can write (69) or (70) in this notation as: $U = [V]$. In differentiating (69) with respect to the coordinates it must be recalled that $[V]$ depends on the coordinates both directly and indirectly via the r, appearing in the fourth argument. We thus get

$$\frac{\partial U}{\partial n} = \left[\frac{\partial V}{\partial n}\right] - \frac{1}{a}\left[\frac{\partial V}{\partial t}\right]\frac{\partial r}{\partial n}. \tag{71}$$

Similarly, on using the expression for Laplace's operator in polar coordinates with centre M_0 [119]:

$$\Delta U = \frac{\partial^2 U}{\partial r^2} + \frac{2}{r}\frac{\partial U}{\partial r} + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial U}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 U}{\partial \varphi^2}$$

and bearing in mind that

$$\frac{\partial U}{\partial \theta} = \left[\frac{\partial V}{\partial \theta}\right]; \quad \frac{\partial^2 U}{\partial \theta^2} = \left[\frac{\partial^2 V}{\partial \theta^2}\right]; \quad \frac{\partial^2 U}{\partial \varphi^2} = \left[\frac{\partial^2 V}{\partial \varphi^2}\right]$$

$$\frac{\partial U}{\partial r} = \left[\frac{\partial V}{\partial r}\right] - \frac{1}{a}\left[\frac{\partial V}{\partial t}\right]; \quad \frac{\partial^2 U}{\partial r^2} = \left[\frac{\partial^2 V}{\partial r^2}\right] - \frac{2}{a}\left[\frac{\partial^2 V}{\partial t\,\partial r}\right] + \frac{1}{a^2}\left[\frac{\partial^2 V}{\partial t^2}\right],$$

we get:

$$\Delta U = [\Delta V] - \frac{2}{a}\left[\frac{\partial^2 V}{\partial t\,\partial r}\right] + \frac{1}{a^2}\left[\frac{\partial^2 V}{\partial t^2}\right] - \frac{2}{ar}\left[\frac{\partial V}{\partial t}\right].$$

But we have, by (68), $[\Delta V] = ,\ 1/a^2\,[\partial^2 V/\partial t^2]$ and therefore:

$$\Delta U = \frac{2}{a}\left\{\frac{1}{a}\left[\frac{\partial^2 V}{\partial t^2}\right] - \left[\frac{\partial^2 V}{\partial t\,\partial r}\right] - \frac{1}{r}\left[\frac{\partial V}{\partial t}\right]\right\}.$$

It may readily be shown that

$$-\frac{\Delta U}{r} = -\frac{2}{a}\left\{\frac{1}{ar}\left[\frac{\partial^2 V}{\partial t^2}\right] - \frac{1}{r}\left[\frac{\partial^2 V}{\partial t\,\partial r}\right] - \frac{1}{r^2}\left[\frac{\partial V}{\partial t}\right]\right\} \tag{72}$$

is the divergence of a vector:

$$-\frac{\Delta U}{r} = \operatorname{div}\left\{\frac{2}{a}\left[\frac{\partial V}{\partial t}\right]\operatorname{grad}(\log r)\right\}. \tag{73}$$

We have in fact by the formula of [112]:

$$\operatorname{div}(f\mathbf{A}) = f\operatorname{div}\mathbf{A} + \operatorname{grad} f \cdot \mathbf{A}.$$

In the present case, $f = 2\,[\partial V/\partial t]/a$, and $\mathbf{A} = \operatorname{grad}(\log r)$ is a vector of length $1/r$ directed along the radius vector from M_0. The scalar product $\operatorname{grad} f.\ \mathbf{A}$ is the product of $|\ \mathbf{A}\ |$ and the projection of $\operatorname{grad} f$ on the direction of $\mathbf{A}$, this latter

term being the derivative of f with respect to the direction of **A**. Thus we have in the present case:

$$\operatorname{div}\left\{\frac{2}{a}\left[\frac{\partial V}{\partial t}\right]\operatorname{grad}(\log r)\right\} = \frac{2}{a}\left[\frac{\partial V}{\partial t}\right]\Delta\log r + \frac{2}{ar}\frac{\partial}{\partial r}\left[\frac{\partial V}{\partial t}\right].$$

If we use (72) and differentiate $[\partial V/\partial t]$ in accordance with the rule for functions of a function, (73) is in fact shown to be valid. If we now apply Ostrogradskii's formula and take into account the fact that $\operatorname{grad}_n(\log r) = (1/r\,\chi\partial r/\partial n)$, we get:

$$-\iiint\limits_{(S)} \frac{\Delta U}{r}\,\mathrm{d}v = \frac{2}{a}\iint\limits_{(S)}\left[\frac{\partial V}{\partial t}\right]\frac{1}{r}\frac{\partial r}{\partial n}\,\mathrm{d}S.$$

On substituting this expression and expression (71) in the right-hand side of (9) and noting that $U(M_0, t) = V(M_0, t)$, since $r = 0$ at the point M_0, we obtain Kirchhoff's formula:

$$V(M_0, t) = \frac{1}{4\pi}\iint\limits_{(S)}\left\{\frac{1}{r}\left[\frac{\partial V}{\partial n}\right] + \frac{1}{ar}\left[\frac{\partial V}{\partial t}\right]\frac{\partial r}{\partial n} - [V]\frac{\partial\frac{1}{r}}{\partial n}\right\}\mathrm{d}S. \tag{74}$$

This expresses $V(M_0, t)$ in terms of the retarded values of V, $\partial V/\partial t$ and $\partial V/\partial n$ on the surface (S). As in the case of (9) for harmonic functions, the presence here of $\partial V/\partial n$ prevents us from using the formula for solving wave equation problems directly. The Kirchhoff formula is closely related to Huygens principle.

Suppose that (S) is a sphere of radius r with centre M_0. Here, $\partial/\partial n = \partial/\partial r$, and (74) becomes:

$$V(M_0, t) = \frac{1}{4\pi r^2}\iint\limits_{(S)}\left\{r\left[\frac{\partial V}{\partial r}\right] + \frac{r}{a}\left[\frac{\partial V}{\partial t}\right] + [V]\right\}\mathrm{d}S,$$

or, on setting $\mathrm{d}S = r^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi = r^2\,\mathrm{d}\omega$:

$$V(M_0;\ t) = \frac{1}{4\pi}\iint\limits_{(S)}\left[\frac{\partial(rV)}{\partial r}\right]\mathrm{d}\omega + \frac{r}{4\pi a}\iint\limits_{(S)}\left[\frac{\partial V}{\partial t}\right]\mathrm{d}\omega. \tag{75}$$

If we take the radius $r = at$, we have $t - r/a = 0$, i.e. the retarded value reduces to the value of the function at $t = 0$, and (75) gives Poisson's formula (76) of [171], i.e. the solution of the problem of an oscillation propagated in unbounded space with given initial conditions:

$$V(M_0, t) = \frac{t}{4\pi}\iint\limits_{(S_{at})}\left(\frac{\partial V}{\partial t}\right)_0\mathrm{d}\omega + \frac{1}{4\pi}\frac{\mathrm{d}}{\mathrm{d}t}\left\{t\iint\limits_{(S_{at})}(V)_0\,\mathrm{d}\omega\right\}, \tag{76}$$

the zero subscript indicating that $\partial V/\partial t$ and V have to be taken at $t = 0$, whilst the integration is over the sphere of radius at with centre M_0. There is a close connection between the form of Kirchhoff's formula and the concept of *retarded potential*. We saw above that, for any choice of $\omega(t)$, having continuous derivatives to the second order, the function

$$\frac{1}{r}\,\omega\left(t - \frac{r}{a}\right) = \frac{[\omega]}{r} \tag{77}$$

is a solution of equation (68). Here, r is the distance from any fixed point of space to a variable point [175].

Precisely as above, we can construct a Kirchhoff formula for any solution of the non-homogeneous wave equation

$$\frac{\partial^2 V}{\partial t^2} = a^2 \Delta V + f(x, y, z, t) \tag{78}$$

in a domain D, though the solution obtained contains a triple as well as surface integral:

$$V(M_0; t) = \frac{1}{4\pi} \iint\limits_S \left\{ \frac{1}{r} \left[\frac{\partial V}{\partial n}\right] + \frac{1}{ar} \left[\frac{\partial V}{\partial t}\right] \frac{\partial r}{\partial n} - [V] \frac{\partial \frac{1}{r}}{\partial n} \right\} dS +$$
$$+ \frac{1}{4\pi a^2} \iiint\limits_D \frac{[f]}{r} dv. \tag{79}$$

On applying this formula to a sphere of radius at with centre M_0 for the solution satisfying zero initial conditions at $t = 0$, we arrive at (91) of [174].

§ 21. The equation of thermal conduction

203. Fundamental equations. We have seen that the equation of thermal conduction in a homogeneous sphere has the form:

$$\frac{\partial U}{\partial t} = a^2 \left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} \right), \tag{1}$$

where

$$a = \sqrt{\frac{k}{\gamma \varrho}}, \tag{2}$$

k is the coefficient of internal conduction, γ is the specific heat of the body and ϱ is the density. In addition to equation (1), we have to bear in mind the *initial conditions*, giving the initial distribution of temperature at $t = 0$:

$$u|_{t=0} = f(x, y, z). \tag{3}$$

If the body has a boundary surface (S), we also have *boundary conditions* on (S), which may vary according to physical factors. For instance, (S) may be held at a definite temperature, which may be variable *with time*. In this case, the boundary conditions imply specifying the function U on (S), this U being possibly dependent on time. If the temperature of the surface is not fixed and there is radiation to the surrounding medium of given temperature U_0, the flow of heat across (S) is proportional by Newton's law (admittedly

far from accurate) to the temperature difference between the medium and the surface (S) of the body. This gives a boundary condition of the form:

$$\frac{\partial U}{\partial n} + h(U - U_0) = 0 \quad (\text{on } S), \tag{4}$$

where the coefficient of proportionality h is the *coefficient of outward thermal conduction.*

For the heat distribution in a body of linear dimensions, i.e. in a homogeneous rod, which we assume to be located along the x axis, we have instead of equation (1):

$$\frac{\partial U}{\partial t} = a^2 \frac{\partial^2 U}{\partial x^2}. \tag{5}$$

This form of equation does not take into account, of course, the heat transfer between the surface of the rod and the surrounding space.

Equation (5) may also be obtained from equation (1) by taking U to be independent of y and z. The initial condition becomes for a rod:

$$U|_{t=0} = f(x). \tag{6}$$

If the rod is finite, we have boundary conditions for both ends. An end can be subject to a given temperature, as above. With radiation, boundary condition (4) takes the form:

$$\frac{\partial U}{\partial x} \mp h(U - U_0) = 0 \quad (\text{at the end}), \tag{7}$$

the $(-)$ sign being used for the left-hand end, with minimum abscissa x, and the $(+)$ sign for the right-hand end, whilst h is a positive constant.

204. Infinite rod. We start with an *infinite rod*, for which we only have to satisfy initial condition (6) in addition to equation (5). We follow Fourier's method and first seek a particular solution of (5) in the form:

$$T(t)\,X(x),$$

which gives us

$$T'(t)\,X(x) = a^2\,T(t)\,X''(x),$$

or

$$\frac{T'(t)}{a^2\,T(t)} = \frac{X''(x)}{X(x)} = -\lambda^2,$$

where λ^2 is a constant. We thus obtain:

$$T'(t) + \lambda^2 a^2 T(t) = 0\,; \quad X''(x) + \lambda^2 X(x) = 0\,, \tag{8}$$

whence, neglecting the constant factor in the expression for $T(t)$:

$$T(t) = \mathrm{e}^{-\lambda^2 a^2 t}, \quad X(x) = A \cos \lambda x + B \sin \lambda x;$$

the constants A and B can depend on λ.

Since we have no boundary condition here, the parameter λ is completely arbitrary, and all values of λ are of the same importance when we form the function $u(x, t)$ as the sum

$$\sum_{(\lambda)} \mathrm{e}^{-\lambda^2 a^2 t} [A(\lambda) \cos \lambda x + B(\lambda) \sin \lambda x]$$

It naturally follows that we replace the *summation* over separate values of λ by the *integral* with respect to λ from $(-\infty)$ to $(+\infty)$, i.e. we put

$$u(x, t) = \int_{-\infty}^{\infty} \mathrm{e}^{-\lambda^2 a^2 t} [A(\lambda) \cos \lambda x + B(\lambda) \sin \lambda x)] \, \mathrm{d}\lambda. \tag{9}$$

We can easily verify that the function written is in fact a solution of equation (5) by using the formula for differentiation under the sign of the definite integral. We now turn to initial condition (6), which gives us

$$u|_{t=0} = f(x) = \int_{-\infty}^{\infty} [A(\lambda) \cos \lambda x + B(\lambda) \sin \lambda x] \, \mathrm{d}\lambda. \tag{10}$$

On comparing the integral on the right with Fourier's formula for the function $f(x)$:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\lambda \int_{-\infty}^{\infty} f(\xi) \cos \lambda (\xi - x) \, \mathrm{d}\xi =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\cos \lambda x \int_{-\infty}^{\infty} f(\xi) \cos \lambda \xi \, \mathrm{d}\xi + \sin \lambda x \int_{-\infty}^{\infty} f(\xi) \sin \lambda \xi \, \mathrm{d}\xi\right] \mathrm{d}\lambda,$$

we see that condition (10) can be satisfied by putting

$$A(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\xi) \cos \lambda \xi \, \mathrm{d}\xi, \quad B(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\xi) \sin \lambda \xi \, \mathrm{d}\xi.$$

Substitution of the expressions obtained for $A(\lambda)$ and $B(\lambda)$ in (9) gives us:

$$u(x,t) = \int_{-\infty}^{\infty} f(\xi)\,d\xi \int_{-\infty}^{\infty} e^{-\lambda^2 a^2 t}[\cos\lambda\xi\cos\lambda x + \sin\lambda\xi\sin\lambda x]\,d\lambda =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\xi)\,d\xi \int_{-\infty}^{\infty} e^{-\lambda^2 a^2 t}\cos\lambda(\xi - x)\,d\lambda =$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} f(\xi)\,d\xi \int_{0}^{\infty} e^{-\lambda^2 a^2 t}\cos\lambda(\xi - x)\,d\lambda, \tag{11}$$

where we have used the fact that the integrand is an even function of λ.

Though the solution of our problem is given by equation (11), this may be simplified on recalling [81] that

$$\int_0^\infty e^{-a^2\lambda^2}\cos\beta\lambda\,d\lambda = \frac{\sqrt{\pi}}{2a}e^{-\frac{\beta^2}{4a^2}},$$

whence

$$\frac{1}{\pi}\int_0^\infty e^{-\lambda^2 a^2 t}\cos\lambda(\xi - x)\,d\lambda = \frac{1}{2a\sqrt{\pi t}}\cdot e^{-\frac{(\xi - x)^2}{4a^2 t}},$$

so that (11) may be written:

$$u(x,t) = \int_{-\infty}^{\infty} f(\xi)\frac{1}{2a\sqrt{\pi t}}e^{-\frac{(\xi - x)^2}{4a^2 t}}\,d\xi. \tag{12}$$

We naturally assume that t is positive in the above and later working. The form of the solution has an important physical significance. We first of all remark that

$$\frac{1}{2a\sqrt{\pi t}}e^{-\frac{(\xi - x)^2}{4a^2 t}}, \tag{13}$$

considered as a function of (x, t), is likewise a solution of equation (5), as is evident from the method by which it is obtained, and as may be verified by direct differentiation. Now what is the physical meaning of this solution?

We distinguish a small element of rod $(x_0 - \delta,\ x_0 + \delta)$ about the point x_0, and let $f(x)$ be zero outside, and a constant U_0 inside the interval $(x_0 - \delta,\ x_0 + \delta)$. This is equivalent physically to com-

municating a quantity of heat $Q = 2\delta c\varrho U_0$ to the element at the initial instant, thus causing a temperature rise of U_0 in the element. The temperature distribution in the rod at subsequent instants is given by equation (12), which becomes in our present case:

$$\int_{x_0-\delta}^{x_0+\delta} U_0 \frac{1}{2a\sqrt{\pi t}} e^{-\frac{(\xi-x)^2}{4a^2t}} d\xi = \frac{Q}{2c\varrho a\sqrt{\pi t}} \frac{1}{2\delta} \int_{x_0-\delta}^{x_0+\delta} e^{-\frac{(\xi-x)^2}{4a^2t}} d\xi.$$

If we now let δ tend to 0, so that the amount of heat Q is distributed over a constantly diminishing section of rod and is communicated in the limit to the point x_0, we arrive at the case of an *instantaneous source of heat at the point* $x = x_0$ *of intensity* Q. The presence of such a source in the rod leads to a temperature distribution given by:

$$\lim_{\delta\to 0} \frac{Q}{2c\varrho a\sqrt{\pi t}} \frac{1}{2\delta} \int_{x_0-\delta}^{x_0+\delta} e^{-\frac{(\xi-x)^2}{4a^2t}} d\xi.$$

Since by the mean value theorem:

$$\frac{1}{2\delta} \int_{x_0-\delta}^{x_0+\delta} e^{-\frac{(\xi-x)^2}{4a^2t}} d\xi = e^{-\frac{(\xi_0-x)^2}{4a^2t}}, \text{ where } x_0 - \delta < \xi_0 < x_0 + \delta,$$

and $\xi_0 \to x_0$ as $\delta \to 0$, the above expression is equal to

$$\frac{Q}{c\varrho} \frac{1}{2a\sqrt{\pi t}} e^{-\frac{(x_0-x)^2}{4a^2t}}$$

It follows that *function* (13) *gives the temperature distribution resulting from an instantaneous heat source of intensity* $Q = c\varrho$ *located at the point* $x = \xi$ *of the rod at the initial instant* $t = 0$ (replacing x_0 by ξ). We can now see the meaning of solution (12). To give the section ξ of the rod a temperature $f(\xi)$ at the initial instant, we have to distribute in the small element $d\xi$ about this point the amount of heat:

$$dQ = c\varrho f(\xi)\, d\xi,$$

or what amounts to the same thing, we have to locate at ξ an instantaneous heat source of intensity dQ; the temperature distribution resulting here is, by (13):

$$f(\xi)\, d\xi \frac{1}{2a\sqrt{\pi t}} e^{-\frac{(\xi-x)^2}{4a^2t}}.$$

The total effect of the initial temperature $f(\xi)$ at all points of the rod is the sum of these separate effects, which gives us solution (12) obtained above:

$$u(x,\,t) = \int_{-\infty}^{\infty} f(\xi)\,\frac{1}{2a\sqrt{\pi t}}\,\mathrm{e}^{-\frac{(\xi-x)^2}{4a^2t}}\,\mathrm{d}\xi.$$

Let the temperature $f(x)$ at the initial instant $t=0$ be zero everywhere outside an interval (a_1, a_2), where it is positive. Solution (12) now becomes:

$$u(x,\,t) = \int_{a_1}^{a_2} f(\xi)\,\frac{1}{2a\sqrt{\pi t}}\,\mathrm{e}^{-\frac{(\xi-x)^2}{4a^2t}}\,\mathrm{d}\xi, \tag{14}$$

If we take an arbitrarily large x and arbitrarily small t, i.e. consider a point as remote, and an instant as near the initial instant, as we wish, (14) gives us a positive value for $u(x, t)$, since the integrand is positive. Solution (12) thus implies the fact that the heat is propagated instantaneously, and not with finite velocity. This is where the thermal conduction equation differs essentially from the wave equation which we obtained when considering the vibrations of rods.

In the case of heat propagation in a non-bounded three-dimensional medium, we have differential equation (1) and initial condition (3), and instead of (12), we get the solution:

$$u(x,\,y,\,z,\,t) =$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(\xi,\,\eta,\,\zeta)\,\frac{1}{(2a\sqrt{\pi t})^3}\,\mathrm{e}^{-\frac{(\xi-x)^2+(\eta-y)^2+(\zeta-z)^2}{4a^2t}}\,\mathrm{d}\xi\,\mathrm{d}\eta\,\mathrm{d}\zeta. \tag{15}$$

We shall verify that the function given by (12) satisfies equation (5) and initial condition (6). The first statement follows immediately from the fact that the integrand of (12) satisfies equation (5) and from the differentiability of the integral of (12) with respect to t and x under the integral sign, if, for instance, $f(x)$ is continuous and absolutely integrable in the interval $(-\infty, +\infty)$. As regards the initial condition, we bring in a new variable α instead of ξ in accordance with

$$\alpha = \frac{\xi - x}{2a\sqrt{t}}\,.$$

Equation (12) now becomes:

$$u(x,\,t) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} f(x + \alpha 2a\sqrt{t})\mathrm{e}^{-\alpha^2}\,\mathrm{d}\alpha. \tag{16}$$

We recall from [78] that

$$1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\alpha^2} \, d\alpha. \tag{17}$$

On multiplying each side by $f(x)$ and subtracting from (16), we have

$$u(x, t) - f(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} [f(x + \alpha 2a \sqrt{t}) - f(x)] \, e^{-\alpha^2} \, d\alpha.$$

whence

$$| u(x, t) - f(x) | \leqslant \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} | f(x + \alpha 2a \sqrt{t}) - f(x) | \, e^{-\alpha^2} \, d\alpha, \tag{18}$$

We shall assume that $f(x)$ is bounded as well as continuous and absolutely integrable, i.e. $| f(x) | \leqslant c$, so that we have for any x, t, and α: $| f(x + \alpha\, 2a \sqrt{t}) - f(x) | \leqslant 2c$. Let ε be a given positive number. We can fix a large positive N, such that

$$\frac{2c}{\sqrt{\pi}} \int_{-\infty}^{-N} e^{-\alpha^2} d\alpha \leqslant \frac{\varepsilon}{3} \text{ and } \frac{2c}{\sqrt{\pi}} \int_{N}^{\infty} e^{-\alpha^2} \, d\alpha \leqslant \frac{\varepsilon}{3}.$$

With this, it follows from (18) that:

$$| u(x, t) - f(x) | \leqslant \frac{2}{3} \varepsilon + \frac{1}{\sqrt{\pi}} \int_{-N}^{+N} | f(x + \alpha 2a \sqrt{t}) - f(x) | \, e^{-\alpha^2} \, d\alpha.$$

Since $f(x)$ is continuous, we can say that, for all t sufficiently near zero and for $| \alpha | \leqslant N$:

$$| f(x + \alpha 2a \sqrt{t}) - f(x) | \leqslant \frac{1}{3} \varepsilon,$$

so that the previous inequality can be written:

$$| u(x, t) - f(x) | \leqslant \frac{2}{3} \varepsilon + \frac{\varepsilon}{3} \cdot \frac{1}{\sqrt{\pi}} \int_{-N}^{+N} e^{-\alpha^2} \, d\alpha,$$

or all the more:

$$| u(x, t) - f(x) | \leqslant \frac{2}{3} \varepsilon + \frac{\varepsilon}{3} \cdot \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\alpha^2} \, d\alpha,$$

i.e. we have by (17): $| u(x, t) - f(x) | \leqslant \varepsilon$ for all t sufficiently near zero. It follows from this, in view of the arbitrariness of ε, that

$$\lim_{t \to +0} u(x, t) = f(x),$$

which amounts to initial condition (6). It may be remarked that t tends to zero through positive values. If m and M are the bounds of $f(x)$, i.e. $m \leqslant f(x) \leqslant M$, it follows from (16) that

$$\frac{m}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\alpha^2} d\alpha \leqslant u(x, t) \leqslant \frac{M}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\alpha^2} d\alpha,$$

Thus we have, by (17), $m \leqslant u(x, t) \leqslant M$, i.e. the temperature $u(x, t)$ lies between the same bounds as does the initial temperature for all positive t. Similar methods may be used to verify expression (15).

205. Semi-infinite rods. Let the rod be bounded at one end, say $x = 0$, so that $x > 0$. We shall assume that there is radiation from the end into the surrounding medium at temperature 0°.

We now have, in addition to initial condition (6), the boundary condition

$$\left.\frac{\partial u}{\partial x}\right|_{x=0} = hu\Big|_{x=0} \tag{19}$$

whereas solution (12) is at once seen to be unsuitable since, in view of the initial condition, $f(x)$ is only defined in $(0, \infty)$. It follows that $f(x)$ must be continued into the interval $(-\infty, 0)$ before we can use (12).

To this end, we rewrite (12) in the form:

$$u(x, t) = \frac{1}{2a\sqrt{\pi t}} \int_0^{\infty} \left[f(\xi) e^{-\frac{(x-\xi)^2}{4a^2 t}} + f(-\xi) e^{-\frac{(x-\xi)^2}{4a^2 t}} \right] d\xi, \tag{20}$$

which is readily justified on splitting $\int\limits_{-\infty}^{+\infty}$ into $\int\limits_{-\infty}^{0}$ and $\int\limits_{0}^{+\infty}$, then replacing ξ in the former by $(-\xi)$. In order to substitute in (19), we evaluate

$$\frac{\partial u}{\partial x} = \frac{1}{2a\sqrt{\pi t}} \int_0^{\infty} \left[f(\xi) \frac{\xi - x}{2a^2 t} e^{-\frac{(x-\xi)^2}{4a^2 t}} - f(-\xi) e \frac{\xi + x}{2a^2 t} e^{-\frac{(x+\xi)^2}{4a^2 t}} \right] d\xi.$$

Hence we deduce that, with $x = 0$:

$$u|_{x=0} = \frac{1}{2a\sqrt{\pi t}} \int_0^{\infty} e^{-\frac{\xi^2}{4a^2 t}} [f(\xi) + f(-\xi)] \, d\xi,$$

$$\left.\frac{\partial u}{\partial x}\right|_{x=0} = \frac{1}{2a\sqrt{\pi t}} \int_0^{\infty} e^{-\frac{\xi^2}{4a^2 t}} [f(\xi) - f(-\xi)] \frac{\xi \, d\xi}{2a^2 t}.$$

Integration by parts gives us:†

$$\int_0^\infty f(\xi)\,e^{-\frac{\xi^2}{4a^2t}}\frac{\xi\,d\xi}{2a^2t} = -\int_0^\infty f(\xi)\,d\left(e^{-\frac{\xi^2}{4a^2t}}\right) = -e^{-\frac{\xi^2}{4a^2t}}f(\xi)\Big|_{\xi=0}^{\xi=\infty}$$

$$+\int_0^\infty f'(\xi)\,e^{-\frac{\xi^2}{4a^2t}}\,d\xi = f(+0)+\int_0^\infty f'(\xi)\,e^{-\frac{\xi^2}{4a^2t}}\,d\xi,$$

and similarly,

$$\int_0^\infty f(-\xi)\,e^{-\frac{\xi^2}{4a^2t}}\frac{\xi\,d\xi}{2a^2t} = f(-0)-\int_0^\infty f'(-\xi)\,e^{-\frac{\xi^2}{4a^2t}}\,d\xi.$$

We assume that $f(x)$ is continued continuously into $(-\infty, 0)$. Clearly, then

$$f(+0)=f(-0)=f(0),$$

and

$$\frac{\partial u}{\partial x}\Big|_{x=0} = \frac{1}{2a\sqrt{\pi t}}\int_0^\infty e^{-\frac{\xi^2}{4a^2t}}\left[f'(\xi)+f'(-\xi)\right]d\xi;$$

condition (19) becomes

$$\frac{1}{2a\sqrt{\pi t}}\int_0^\infty e^{-\frac{\xi^2}{4a^2t}}\left\{\left[f'(\xi)+f'(-\xi)\right]-h\left[f(\xi)+f(-\xi)\right]\right\}d\xi=0,$$

which is certainly satisfied if we put

$$f'(-\xi)+f'(\xi)=h\left[f(-\xi)+f(\xi)\right],$$

or, using for the present the notation

$$\Phi(\xi)=f(-\xi);\quad \Phi'(\xi)=-f'(-\xi),$$

if we determine the unknown function $\Phi(\xi)$ from the differential equation

$$\Phi'(\xi)+h\,\Phi(\xi)=f'(\xi)-h\,f(\xi)\qquad(\xi>0),$$

We find by integrating this equation:

$$\Phi(\xi)=e^{-h\xi}\left\{C+\int_0^\xi e^{h\xi}\left[f'(\xi)-h\,f(\xi)\right]d\xi\right\}.$$

We find the constant C by setting $\xi=0$:

$$C=\Phi(0)=f(0),$$

† We assume that $e^{-\xi^2/4a^2t}f(\xi)\to 0$ as $\xi\to\infty$.

and since

$$\int_0^\xi e^{h\xi} f'(\xi)\,d\xi = f(\xi)\,e^{h\xi}\Big|_{\xi=0}^{\xi=\xi} - h\int_0^\xi e^{h\xi} f(\xi)\,d\xi = e^{h\xi} f(\xi) - f(0) - h\int_0^\xi e^{h\xi} f(\xi)\,d\xi\,,$$

we have

$$f(-\xi) = \Phi(\xi) = f(\xi) - 2he^{-h\xi}\int_0^\xi e^{h\xi} f(\xi)\,d\xi.$$

Substitution of this expression for $f(-\xi)$ in (20) gives us the final solution of our problem. We remark that it follows from the last expression that, as $\xi \to +0$, $f(-0) = f(+0)$, i.e. $f(x)$ is continued continuously into $(-\infty, 0)$, as we assumed above.

If, for example, the initial temperature is constant:

$$f(x) = u_0 \qquad x > 0\,,$$

we have

$$f(-x) = u_0 - 2he^{-hx}\int_0^x u_0\,e^{hx}\,dx = u_0\,(2e^{-hx} - 1)\,,$$

and (20) gives

$$u(x,t) = \frac{u_0}{2a\sqrt{\pi t}}\left\{\int_0^\infty e^{-\frac{(\xi-x)^2}{4a^2t}}\,d\xi - \int_0^\infty e^{-\frac{(\xi+x)^2}{4a^2t}}\,d\xi + \right.$$

$$\left. + 2\int_0^\infty e^{-\frac{(\xi+x)^2}{4a^2t} - h\xi}\,d\xi\right\}. \tag{21}$$

It may readily be shown by the reader that this solution can be expressed in terms of the function

$$\Theta(x) = \frac{2}{\sqrt{\pi}}\int_0^\infty e^{-x^2}\,dx$$

as follows:

$$u(x,t) = u_0\Theta\left(\frac{x}{2a\sqrt{t}}\right) + u_0\,e^{a^2h^2t+hx}\left[1 - \Theta\left(\frac{x}{2a\sqrt{t}} + ah\sqrt{t}\right)\right]. \tag{22}$$

A simpler result is obtained in the case when there is no radiation at the end $x = 0$, this being held at a temperature of 0°. We now have the boundary condition

$$u|_{x=0} = 0\,, \tag{23}$$

which can be obtained from (19) by dividing by h then passing to the limit with $h \to \infty$. The solution can be found from (22) by letting $h \to \infty$, but a simpler method is to continue $f(x)$ directly into $(-\infty, 0)$ in such a way as to satisfy

$$u|_{x=0} = \frac{1}{2a\sqrt{\pi t}}\int_0^\infty e^{-\frac{\xi^2}{4a^2t}}\,[f(\xi) + f(-\xi)]\,d\xi = 0\,,$$

which simply means putting

$$f(-\xi) = -f(\xi),$$

i.e. continuing $f(x)$ in an odd manner.

We now obtain instead of (20):

$$u(x, t) = \frac{1}{2a\sqrt{\pi t}} \int_0^\infty f(\xi)\left[e^{-\frac{(x-\xi)^2}{4a^2t}} - e^{-\frac{(x+\xi)^2}{4a^2t}}\right] d\xi, \tag{24}$$

and if

$$u|_{t=0} = f(x) = u_0.$$

the solution becomes

$$u(x, t) = \frac{u_0}{\sqrt{\pi}} \int_{-\frac{x}{2a\sqrt{t}}}^{\frac{x}{2a\sqrt{t}}} e^{-\xi^2} d\xi = u_0\, \Theta\left(\frac{x}{2a\sqrt{t}}\right). \tag{25}$$

We now consider a rod bounded at the end $x = 0$ and subjected to a given temperature $u = \varphi(t)$ at this end.

We first suppose that the initial temperature is $0°$, i.e.

$$u|_{t=0} = 0, \tag{26}$$

and we start with the particular case $\varphi(t) = 1$, i.e.

$$u|_{x=0} = 1. \tag{27}$$

We can easily find the solution of equation (5) satisfying conditions (26) and (27). We do this by setting

$$u = v + 1;$$

function v is likewise a solution of (5), but has to satisfy the conditions

$$v|_{x=0} = 0; \quad v_{t=0} = -1,$$

so that v and therefore u is obtained directly from (25) by putting $u_0 = -1$ there:

$$v(x, t) = -\Theta\left(\frac{x}{2a\sqrt{t}}\right) \quad \text{and} \quad u(x, t) = 1 - \Theta\left(\frac{x}{2a\sqrt{t}}\right). \tag{28}$$

We now find the temperature distribution if the end $x = 0$ is held at $0°$ up to the instant τ then raised to $1°$. Let the distribution be denoted by $u_\tau(x, t)$. We obviously have $u_\tau = 0$ to the instant $t = \tau$; thereafter u_τ is the same as the solution found above if we agree to reckon t from τ instead of 0, i.e. if we put $t - \tau$ for t in (28); this gives us

$$u_\tau(x, t) = \begin{cases} 0 & \text{for} \quad t < \tau \\ 1 - \Theta\left(\dfrac{x}{2a\sqrt{t-\tau}}\right) & \quad t > \tau. \end{cases}$$

But it is now clear that, if the end $x = 0$ is held at a temperature of $1°$ only during the interval $(\tau, \tau + d\tau)$ whilst being $0°$ throughout the rest of the time, the corresponding distribution becomes

$$u_\tau(x, t) - u_{\tau + d\tau}(x, t) = -\frac{\partial u_\tau}{\partial \tau} d\tau ;$$

whereas if it were subjected to a temperature $\varphi(\tau)$ instead of $1°$ during the interval $(\tau, \tau + d\tau)$, the solution would be

$$-\varphi(\tau)\frac{\partial u_\tau}{\partial \tau} d\tau ,$$

hence it is evident that, if the end $x = 0$ is subject to the temperature $\varphi(\tau)$ for all $\tau > 0$, we obtain the full effect as τ varies from 0 to t by summing all the elementary effects, which gives us the required solution in the form

$$u(x, t) = -\int_0^t \varphi(\tau)\frac{\partial u_\tau}{\partial \tau} d\tau ,$$

or, since we have for $t > \tau$:

$$-\frac{\partial u_\tau}{\partial \tau} = \frac{\partial}{\partial \tau}\Theta\left(\frac{x}{2a\sqrt{t-\tau}}\right) = \frac{\partial}{\partial \tau}\frac{2}{\sqrt{\pi}}\int_0^{\frac{x}{2a\sqrt{t-\tau}}} e^{-x^2} dx = \frac{x}{2a\sqrt{\pi}(t-\tau)^{3/2}} e^{-\frac{x^2}{4a^2(t-\tau)}},$$

the solution is finally

$$u(x, t) = \frac{x}{2a\sqrt{\pi}}\int_0^t \frac{\varphi(\tau)}{(t-\tau)^{3/2}} e^{-\frac{x^2}{4a^2(t-\tau)}} d\tau. \tag{29}$$

We evidently get the solution which satisfies, in addition to the boundary condition

$$u|_{x=0} = \varphi(t)$$

an initial condition of the general type

$$u|_{t=0} = f(x) ,$$

[nstead of the particular condition (26)] simply by adding to solution (29) the solution that we previously obtained, given by (24).

206. Rods bounded at both ends. We examine one of the most typical cases, when the temperature at the end $x = 0$ is $0°$:

$$u|_{x=0} = 0 ; \tag{30}$$

whilst heat is radiated from the end $x = l$ into the surrounding medium with zero temperature:

$$\left.\frac{\partial u}{\partial x}\right|_{x=l} = -hu\Big|_{x=l} ; \tag{31}$$

and the initial temperature is

$$u\big|_{t=0} = f(x) \quad (0 \leqslant x \leqslant l). \tag{32}$$

This problem is solved very simply by Fourier's method.

Since boundary conditions are present here, we subject the solution found above:

$$e^{-\lambda^2 a^2 t} X(x) = e^{-\lambda^2 a^2 t}[A \cos \lambda x + B \sin \lambda x] \tag{33}$$

to conditions (30) and (31), which give us

$$X(0) = 0, \quad \text{i.e.} \quad A = 0; \quad X'(l) = -hX(l),$$

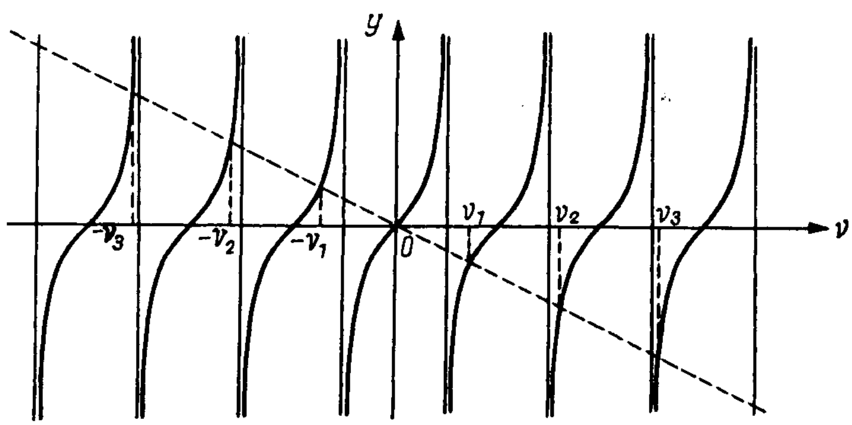

Fig. 143

whence we obtain, neglecting the constant factor B:

$$X(x) = \sin \lambda x \tag{34}$$

and

$$\lambda \cos \lambda l = -h \sin \lambda l. \tag{35}$$

On substituting $\lambda l = v$, we obtain the transcendental equation

$$\tan v = \alpha v, \text{ where } \alpha = -\frac{1}{hl}. \tag{36}$$

This equation has an infinite set of real roots (Fig. 143). We shall only take into account the positive roots:

$$v_1, v_2, v_3 \ldots, v_n, \ldots \tag{37}$$

corresponding to which we have an infinite set of values of λ:

$$\lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_n, \ldots, \text{ where } \lambda_n = \frac{v_n}{l}, \tag{38}$$

and to which, in turn, there corresponds an infinite set of particular solutions of equation (5):

$$B_n \mathrm{e}^{-\lambda_n^2 a^2 t} \sin \lambda_n x \quad (x = 1, 2, 3, \dots),$$

satisfying the boundary conditions.

In order to satisfy the initial conditions, we seek u as

$$u(x, t) = \sum_{n=1}^{\infty} B_n \mathrm{e}^{-\lambda_n^2 a^2 t} \sin \lambda_n x, \tag{39}$$

and obtain with $t = 0$:

$$u|_{t=0} = f(x) = \sum_{n=1}^{\infty} B_n \sin \lambda_n x = \sum_{n=1}^{\infty} B_n X_n(x), \tag{40}$$

where we have written $X_n(x) = \sin \lambda_n x$. We shall prove that the functions $X_n(x)$ are orthogonal.

We write down the differential equations corresponding to two of these functions:

$$X_m''(x) + \lambda_m^2 X_m(x) = 0; \quad X_n''(x) + \lambda_n^2 X_n(x) = 0.$$

On multiplying the first equation by $X_n(x)$ and the second by $X_m(x)$, subtracting the equations obtained and integrating over the interval $(0, l)$, we have:

$$\int_0^l \left[X_m''(x) X_n(x) - X_n''(x) X_m(x)\right] \mathrm{d}x + \\ + \left(\lambda_m^2 - \lambda_n^2\right) \int_0^l X_m(x) X_n(x)\, \mathrm{d}x = 0.$$

On integrating the first integral by parts, we get

$$X_m'(l) X_n(l) - X_n'(l) X_m(l) + X_n'(0) X_m(0) - X_m'(0) X_n(0) + \\ + (\lambda_m^2 - \lambda_n^2) \int_0^l X_m(x) X_n(x)\, \mathrm{d}x = 0. \tag{41}$$

But $X_m(x)$ and $X_n(x)$ satisfy boundary conditions (30) and (31), i.e.

$$X_m(0) = X_n(0) = 0; \quad X_m'(l) = -hX_m(l);$$
$$X_n'(l) = -hX_n(l).$$

In view of these equations, the term outside the integral in (41) vanishes, and if we bear in mind that $\lambda_m^2 - \lambda_n^2 \neq 0$ for different m and n, we get

$$\int_0^l X_m(x)\, X_n(x)\, dx = 0 \quad \text{for} \quad m \neq n.$$

Having established orthogonality, we proceed as usual to show that the coefficients B_n in expansion (40) are given by:

$$B_n = \int_0^l f(x)\, X_n(x)\, dx : \int_0^l X_n^2(x)\, dx.$$

This solves the problem of expanding $f(x)$ in functions $X_n(x)$ and at the same time gives the solution of our main problem in the form of series (39). It will be shown in Vol. IV that the $X_n(x)$ obtained, as above, as a result of applying Fourier's method to typical problems of mathematical physics form a closed system, and that, with certain assumptions regarding $f(x)$, this latter can be expanded in the base interval as a uniformly convergent series in $X_n(x)$. It may be noted that, if we took boundary conditions $u = 0$ for $x = 0$ and $x = l$, instead of (30) and (31), we should get $X_n(x) = \sin n\pi x/l$, and arrive at the usual Fourier sine expansion.

When investigating the propagation of heat in a ring, in addition to the boundary conditions, we have to stipulate periodicity of the temperature [cf. 195]. If we take the radius of the ring as unity, so that its total length is 2π, and let x denote the length along the ring measured from a given point, we arrive at a solution of the form

$$u(x, t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)\, e^{-a^2 n t},$$

where

$$\frac{a_0}{2} + \sum_{n=1}^{a} (a_n \cos nx + b_n \sin nx)$$

is the Fourier series for the initial temperature distribution $f(x)$ on the ring.

Sufficient conditions for the series obtained for $u(x, t)$ to be the actual solution of our problem will be found in Vol. IV.

207. Supplementary remarks. We take the general equation for thermal conduction:

$$\frac{\partial v}{\partial t} = a^2 \frac{\partial^2 v}{\partial x^2} - cv, \tag{42}$$

obtained on the assumption of radiation from the total surface of the rod to the surrounding medium, the temperature of which is taken as zero.

We can easily verify that equation (42) reduces to equation (5) for u by the simple substitution

$$v = \mathrm{e}^{-ct} u.$$

In the case of an infinite rod with zero initial temperature, i.e. with the condition $u = 0$ for $t = 0$, the non-homogeneous equation

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + F(x, t) \tag{43}$$

has a solution of the form

$$u(x, t) = \int_0^t \int_{-\infty}^{+\infty} F(\xi, \tau) \frac{1}{2a\sqrt{\pi(t-\tau)}} \mathrm{e}^{-\frac{(\xi-x)^2}{4a^2(t-\tau)}} \mathrm{d}\xi \, \mathrm{d}\tau. \tag{44}$$

It can be found either by the same method as was used in [174] for the non-homogeneous wave equation, or by superposition of the basic elementary solution (13), in which we replace t by $(t - \tau)$ then multiply by $F(\xi, \tau)$ and integrate, from $-\infty$ to $+\infty$ with respect to ξ, and from 0 to t with respect to τ. The physical significance of these operations is obvious. The solution of (43) is found by means of superposition of sources, distributed throughout the rod with intensity $F(\xi, \tau)$, and starting to be effective at the instant τ. Superposition of the sources is also carried out with respect to time.

Fourier's method is applied to two and three-dimensional cases exactly as for the wave equation, except that the time-dependent factor is now an exponential function.

For instance, the equation

$$\frac{\partial u}{\partial t} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right)$$

leads in the case of a rectangular lamina to a solution of the form

$$u = \mathrm{e}^{-\omega^2 t} U(x, y), \tag{45}$$

where we have written ω^2 in the exponent so as to be able to use the expressions of [177]. Let the boundary condition be $u = 0$ on C and the initial condition $u = \varphi_1(x, y)$ for $t = 0$. The solution takes the form of the series:

$$u = \sum_{\sigma,\tau=1}^{\infty} a_{\sigma,\tau} \mathrm{e}^{-\omega^2_{\sigma,\tau} t} \sin \frac{\sigma \pi x}{\mathrm{e}} \sin \frac{\tau \pi y}{m},$$

where the $\omega^2_{\sigma,\tau}$ are given by (113) of [177], and the $a_{\sigma,\tau}$ by the first of formulae (114).

In the case of a circular lamina [cf. 178], substitution (45) leads to the following solution:

$$u = \sum_{\substack{n=0\\m=1}}^{\infty} a_{m,n} \mathrm{e}^{-\omega^2_{m.n} t} \cos n\theta J_n (k_m^{(n)} r) + \sum_{\substack{n=1\\m=1}}^{\infty} \beta_{m,n} \mathrm{e}^{-\omega^2_{m.n} t} \sin n\theta J_n (k_m^{(n)} r),$$

where the $\alpha_{m,n}$ and $\beta_{m,n}$ are given by the same formulae as the $\alpha^{(1)}_{m,n}$ and $\beta^{(1)}_{m,n}$ of [178], and the $\omega_{m,n}$ by (128).

208. The case of a sphere. We give a parallel treatment of the wave equation and thermal conduction equation for a sphere:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \Delta u ; \tag{46}$$

$$\frac{\partial v}{\partial t} = a^2 \Delta v , \tag{47}$$

with the assumption that the initial data depend only on the distance r of a point from the centre of the sphere:

$$u\Big|_{t=0} = \varphi_1(r) ; \quad \frac{\partial u}{\partial t}\Big|_{t=0} = \varphi_2(r) ; \tag{48}$$

$$v|_{t=0} = \psi(r). \tag{49}$$

We take the boundary conditions:

$$\frac{\partial u}{\partial r} = 0 \qquad r = R ; \tag{50}$$

$$\frac{\partial v}{\partial r} + hv = 0 \qquad r = R , \tag{51}$$

where R is the radius of the sphere and $h > 0$. In view of the central symmetry, the solution will be independent of the polar angle and will be a function of r and t only. On setting:

$$u = (A\cos\omega t + B\sin\omega t)\, U(r) ; \tag{52}$$

$$v = A e^{-\omega^2 t}\, V(r) , \tag{53}$$

we obtain the same equation $\Delta W + k^2 W = 0$ for $U(r)$ and $V(r)$, where $k^2 = \omega^2/a^2$. On writing Laplace's operator in spherical coordinates and using the fact that W depends only on r, we get:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2 \frac{dW}{dr}\right) + k^2 W = 0 , \text{ i.e. } \frac{d^2 W}{dr^2} + \frac{2}{r}\frac{dW}{dr} + k^2 W = 0 .$$

We bring in a new unknown instead of W, given by

$$R(r) = rW(r).$$

On substituting in the equation for W, we get the equation for $R : R''(r) + k^2 R(r) = 0$, whence $R(r) = C_1 \cos kr + C_2 \sin kr$, and consequently:

$$W(r) = C_1 \frac{\cos kr}{r} + C_2 \frac{\sin kr}{r} .$$

In view of the solution remaining finite at the centre of the sphere, i.e. with $r = 0$, we must write $C_1 = 0$, and substitution in (52) gives us the solution as

$$u = (A \cos \omega t + B \sin \omega t) \frac{\sin kr}{r}; \tag{54}$$

$$v = A e^{-\omega^2 t} \frac{\sin kr}{r}. \tag{55}$$

Boundary conditions (50) and (51) give us the constant k, and consequently, $\omega = ak$.

Application of the second of these to $(\sin kr)/r$ results in the following equation for k:

$$kR \cot kR = 1 - hR. \tag{56}$$

With $h = 0$, we arrive at the equation obtained from boundary condition (50):

$$\tan kR = kR. \tag{57}$$

On putting $kR = v$, we see that equations (56) and (57) are analogous to equation (36). Let $k_1, k_2, \ldots$ be the positive roots of (56). On using (55), we get for $v(r, t)$:

$$v(r, t) = \sum_{n=1}^{\infty} a_n e^{-a^2 k_n^2 t} \frac{\sin k_n r}{r}. \tag{58}$$

Initial condition (49) gives:

$$r\psi(r) = \sum_{n=1}^{\infty} a_n \sin k_n r. \tag{59}$$

Precisely as in [206], the functions $\sin k_n r$ are orthogonal in the interval $(0, R)$, and the coefficients of expansion (59) are therefore given by:

$$a_n = \int_0^R r\psi(r) \sin k_n r \, dr : \int_0^R \sin^2 k_n r \, dr.$$

Turning to the equation for u, we first write k_n ($n = 1, 2, \ldots$) for the positive roots of equation (57). We have to include here the root $k = 0$, corresponding to a zero frequency ω. With this, we have to write $A + Bt$ instead of $(A \cos \omega t + B \sin \omega t)$, the equation for $R(r)$ becomes $R''(r) = 0$, and $W(r) = R(r)/r$ is constant, so that the corresponding solution of equation (46) becomes $a_0 + b_0 t$. This clearly satisfies boundary condition (50) for any values of constants a_0 and b_0. We finally get for u:

$$u(r, t) = a_0 + b_0 t + \sum_{n=1}^{\infty} (a_n \cos ak_n t + b_n \sin k_n t) \frac{\sin k_n r}{r}.$$

On differentiating with respect to t and setting $t = 0$, we get the expansions for the functions appearing in initial conditiosu (48):

$$r\varphi_1(r) = a_0 r + \sum_{n=1}^{\infty} a_n \sin k_n r; \quad r\varphi_2(r) = b_0 r + \sum_{n=1}^{\infty} k_n b_n \sin k_n r.$$

It may easily be seen by using (57) that the $\sin k_n r$ are orthogonal to r in the interval $(0, R)$ as well as to each other, i.e.

$$\int_0^R r \sin k_n r \, dr = 0 \quad \text{and} \quad \int_0^R \sin k_m r \sin k_n r \, dr = 0\,, \quad \text{for} \quad m \neq n\,,$$

and the coefficients in the last expansions are given by the usual rule:

$$a_0 = \int_0^R r^2 \varphi_1(r) dr : \int_0^R r^2 \, dr = \frac{3}{R^3} \int_0^R r^2 \varphi_1(r) \, dr\,,$$

$$a_n = \int_0^R r \varphi_1(r) \sin k_n r \, dr : \int_0^R \sin^2 k_n r \, dr.$$

Similar expressions are found for coefficients b_n. We remark that the solution $v = \text{const.}$ is obtained for equation (47) with $\omega = 0$, but this solution does not satisfy boundary condition (51), since $h > 0$ by hypothesis.

We can interpret (46) as the equation for the velocity potential u of the vibrations of a gas, where boundary condition (50) expresses the fact that the velocity of a gaseous particle situated on the surface of the sphere has a zero normal component.

Boundary condition (51) for the thermal conduction equation (47) expresses the fact that the surface of the sphere radiates to the surrounding space, the temperature of which is held at zero.

209. The uniqueness theorem. We now pass to the question of the uniqueness of the solution of the thermal conduction equation with given initial and boundary conditions [cf. 179]. We take the one-dimensional problem, i.e. the equation:

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} \tag{60}$$

for a finite rod, $0 \leqslant x \leqslant l$. We draw a domain G on the xt plane, bounded by the lines $x = 0$, $x = l$, and situated above the segment $0 \leqslant x \leqslant l$ of the x axis (Fig. 144). We also draw a line $t = t_0$, $t_0 > 0$, parallel to the x axis, and cutting a finite rectangle $OAQP$ from the domain G, the rectangle being denoted simply by H. We prove the following theorem:

THEOREM. *Let the function $u(x, t)$ satisfy equation* (60) *inside G and be continuous as far as the contour of G. The greatest and least values of $u(x, t)$ in H are now attained on the part l of the contour of H, formed by sides OP, PA and AQ.*

We confine the proof to the greatest value and assume the theorem false. Let the greatest value of $u(x, t)$ be attained in H or within the

side PQ at the point (x, y'), so that the greatest value on l is less than this first greatest value M. We construct a new function $v(x, t)$ as follows:

$$v(x, t) = u(x, t) - k\ (t - t_0), \tag{61}$$

where k is a positive number which we now fix. We have in the rectangle H:

$$u(x, t) \leqslant v(x, t) \leqslant u(x, t) + kt_0,$$

and k can be fixed near enough zero for the greatest value of $v(x, t)$ on l to be, as in the case of $u(x, t)$ also, less than the value of $v(x, t)$

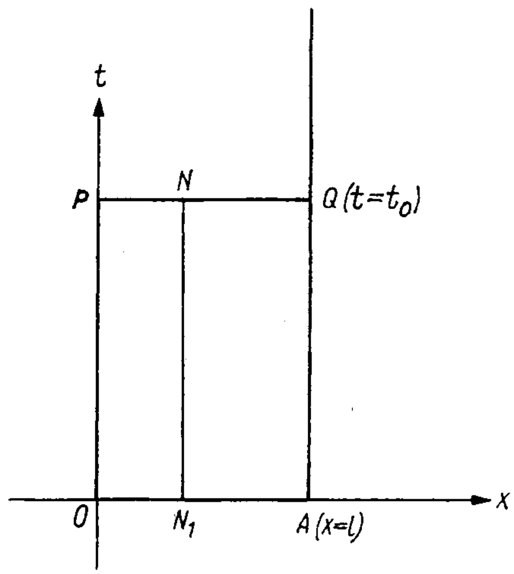

FIG. 144

at the point (x', t'). With this choice of k, $v(x, t)$ will take its greatest value in H inside H or within the side PQ, and not on l. We consider these cases separately and prove that each leads to a contradiction.

Let $v(x, t)$ take its greatest value at a point $C(x_1, t_1)$ situated inside H. The fact that we have a maximum of $v(x, t)$ at C implies that, at this point [I, 58]:

$$\frac{\partial v}{\partial t} = 0 \quad \text{and} \quad \frac{\partial^2 v}{\partial x^2} \leqslant 0,$$

whence it follows that

$$\frac{\partial v}{\partial t} - a^2 \frac{\partial^2 v}{\partial x^2} \geqslant 0,$$

or, by (61):

$$\frac{\partial u}{\partial t} - a^2 \frac{\partial^2 u}{\partial x^2} - k \geqslant 0.$$

But u satisfies equation (60) inside G, and the inequality written leads to the absurd result: $-k \geqslant 0$. We now let $v(x, t)$ attain its greatest value in H at a point $N(x_1, t_0)$ situated within the side PQ. On considering the variation of $v(x, t)$ along $N_1 N$, parallel to the t axis, we arrive at the inequality $\partial v/\partial t \geqslant 0$ at N, inasmuch as the value of $v(x, t)$ at N is not less than its values throughout $N_1 N$. By now considering the variation of $v(x, t)$ along PQ, we arrive at the inequality $\partial^2 v/\partial x^2 \leqslant 0$ at the point N, since $v(x, t_0)$ has a maximum at N $(x = x_1)$. Thus $\partial v/\partial t - a^2 \partial^2 v/\partial x^2 \geqslant 0$ at N, and we arrive at the same contradiction as above, so that the theorem is proved.

It immediately follows from this theorem that, if $u(x, t)$ vanishes all along the contour l, $u(x, t)$ also vanishes throughout the rectangle H, which leads us very simply to the uniqueness theorem.

Suppose that, in addition to equation (60), we have the following initial and boundary conditions (specifying the temperature at the ends):

$$u\big|_{t=0} = f(x) \ (0 \leqslant x \leqslant l)\,; \quad u\big|_{x=0} = \omega(t)\,; \quad u\big|_{x=l} = \omega_1(t). \qquad (62)$$

These conditions amount to specifying $u(x, t)$ on the piece l of the contour of G. We assume that these boundary values represent functions continuous throughout the contour of G, including points O and A, i.e. $\omega(0) = f(0)$ and $\omega_1(0) = f(l)$. With conditions (62), let there exist inside G two solutions of equation (60), $u_1(x, t)$ and $u_2(x, t)$, continuous as far as the contour of G. Their difference, $u(x, t) = u_1(x, t) - u_2(x, t)$, is now a solution of (60), equal to zero on l. It follows at once from the theorem proved above that u is zero everywhere inside G, i.e. $u_1(x, t)$ coincides with $u_2(x, t)$. We remark that the uniqueness theorem is preserved if, instead of demanding continuity of $u(x, t)$ at points O and A, we only require boundedness of the function in the neighbourhood of these points. In this case, the boundary values likewise do not need to be continuous at O and A.

The solution for an infinite rod is given by (12). We suppose that the given function $f(x)$ is continuous and vanishes outside the segment $(-b, +b)$, so that

$$u(x, t) = \frac{1}{2a\sqrt{\pi t}} \int_{-b}^{+b} f(\xi)\, e^{-\frac{(\xi - x)^2}{4a^2 t}}\, d\xi.$$

It is easily shown by using this formula that $u(x, t) \to 0$ uniformly with respect to t as $x \to +\infty$ or $x \to -\infty$, i.e. for any given positive ε

there exists a positive N such that $|u(x, t)| \leqslant \varepsilon$ for $|x| \geqslant N$ and for any positive t. We prove that only one solution exists with this property for a given initial condition (6). It is sufficient to show, as above, that $u(x, t)$ takes its greatest and least values on the x axis. We do this by *reductio ad absurdum*. Let $u(x, t)$ take its greatest value M at a point $C(x_1, t_1)$, where $t_1 > 0$, i.e. $f(x) < M$ in the interval

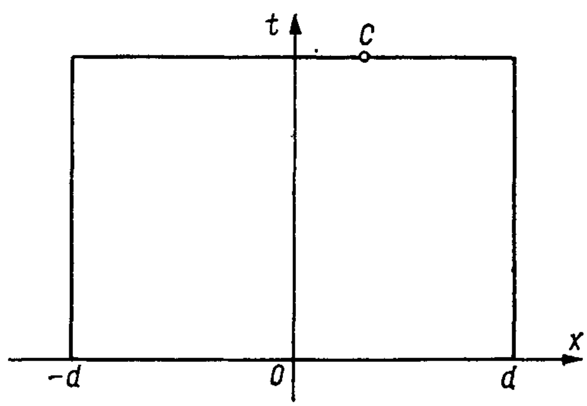

FIG. 145

$-\infty < x < +\infty$. In view of the fact that $f(x) = 0$ outside the interva $(-b, b)$, we can say that $M > 0$. We draw the two lines $x = d$ and $x = -d$, choosing d sufficiently large for the inequality $|u(x, t)| < M$ to be valid on these lines, then we draw the rectangle H formed by these lines, the x axis, and the straight line through C parallel to the x axis (Fig. 145). The function $u(x, t)$ has a greater value at C than on the piece l of the contour of H, consisting of the three sides: $x = d$, $x = -d$ and $t = 0$. Thus $u(x, t)$ attains its greatest value in regard to H either inside H or within the side passing through C, and this leads to a contradiction, as above. Hence we have proved the uniqueness of the solution having the property stated above and with the assumptions made regarding $f(x)$.

INDEX